Neuro-cognitive Processes Implicated in Passive Avoidance Learning, Probabilistic Reversal Learning and the Development of Psychopathic Tendencies

Salima Budhani

Department of Psychology, University College London

Thesis submitted for the degree of Ph.D. 2005

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgements

I would like to thank James Blair for his indispensable supervision and guidance. I would also like to thank all members of James' laboratory, past and present, who have provided help and advice. Thanks especially go to Derek Mitchell, Rebecca Richell, Karina Peschardt, Marina Nakic and Essi Viding. I am extremely grateful to Ziad Saad and Gang Chen for statistical help, Val Curran for supervision, Danny Pine for comments and suggestions and John Draper for all of his administrative help over the past 4 years. I would also like to thank my family and friends who have encouraged and supported me throughout this process, thanks especially must go to Matthew.

This Ph.D. was supported by a Medical Research Council studentship.

Abstract

Chapter 1 introduced psychopathy and evaluated theories attempting to explain this disorder. Experiment 1 assessed passive avoidance learning in children with psychopathic tendencies. Results replicated previous findings with psychopathic adults indicating that this disorder is associated with poor passive avoidance learning. Experiment 2 developed a connectionist model of passive avoidance learning, the output of which was compared with the results obtained in experiment 1. The intact model successfully simulated performance of the comparison children whilst a model impaired in the formation of stimulus-punishment associations most successfully captured the performance of the children with psychopathic tendencies. Experiment 3 assessed the Blood Oxygen Level-Dependent (BOLD) responses associated with passive avoidance learning in healthy adults. Results revealed that successful passive avoidance learning was associated with activation within rostral anterior cingulate cortex, insula, caudate, hippocampus, and the amygdala. Experiment 4 assessed the performance of children with psychopathic tendencies on a novel probabilistic reversal learning paradigm. Results revealed that children with psychopathic tendencies presented with impairment only on the probabilistic contingencies. Further, it was revealed that the children with psychopathic tendencies committed more win-shift errors in the reversal phases. Experiment 5 assessed the performance of adult individuals with psychopathy using a similar task. Results revealed that adults with psychopathy were impaired in both the simple and probabilistic conditions. The adults with psychopathy also committed more win-shift responses in the reversal phases. Experiment 6 assessed the BOLD responses associated with probabilistic reversal learning in healthy adults. Results revealed that errors in both acquisition and reversal phases were associated with activations within dorsomedial and ventrolateral PFC and caudate, and deactivations within medial OFC cortex, amygdala and hippocampus. Chapter 6 re-evaluated the theories of psychopathy in light of the empirical work presented in this thesis, and discussed the implications of these results along with future research directions.

Table of Contents

N.B. Some experiments within this thesis have been published, are currently in press, or are under review in peer-reviewed journals. Specifically, experiment 3 is in press at *Neuroimage*; experiment 4 is in press at *The Journal of Abnormal Psychology*; experiment 5 is published within *The Journal of Child Psychology and Psychiatry*; and experiment 6 is under review at *Neuroimage*.

List of Tables

List of Figures

Chapter 1 - Introduction

This chapter will introduce the phenomenon of psychopathy. Specifically research will be presented indicating that a small subset of antisocial individuals exist, who begin their criminal careers early in life and engage in a disproportionate amount of antisocial, violent and criminal behaviours. It will be suggested that at least some of these individuals are presenting with the disorder of psychopathy. Current tools for the measurement of psychopathy will be described and the important issue of co-morbidity between psychopathic tendencies and attention-deficit/hyperactivity disorder in childhood will be broadly discussed. The second half of this chapter will consider theories attempting to explain psychopathy.

1.1: Introduction to Antisocial Behaviour

There is considerable concern about the level of antisocial behaviour in modern societies. A recent Home Office report estimated that over 13 million crimes were committed against adults in England and Wales in 2001; approximately 3 million of these were violent crimes (Simmons, 2002). Further, in the UK the number of young people committing 'grave' crimes, such as murder, manslaughter and grievous bodily harm, has almost doubled in recent years (Simmons, 2002). Indeed, since 2001, rates of serious violent crimes against the person have increased by 15% (Dodd et al., 2004). Such high levels of crime and antisociality incur a huge financial cost to society. Scott and colleagues (2001) followed a group of 47 antisocial children into adulthood and observed that these children cost an average of 10 times more than children without behavioural disorders, with crime incurring the greatest cost. The prevalence and costs of violence in our society has stimulated both social and biological scientists to search for the predictors and causes of this destructive human behaviour.

In healthy community samples, during adolescence, antisocial behaviour can be normative (Moffitt, 1993a). By early adulthood, however, the number of active offenders decreases by 50%, and by age 28 over 85% of former offenders

have usually desisted (Blumstein and Cohen, 1987; Farrington, 1986). Moffitt (Moffitt, 1993a) coined the term 'adolescence-limited' to describe this group of offenders whose antisocial behaviour is confined almost exclusively to adolescence. Research has also identified a group whose antisocial behaviour does not cease in adolescence, but instead continues into adulthood. For this group Moffitt coined the term 'life-course persistent' offenders (Moffitt, 1993a). Notably this group also tends to begin their criminal careers prior to adolescence. In keeping with this, studies have revealed that antisocial behaviour rarely presents for the first time in adulthood (*i.e.* antisocial adults were usually antisocial during childhood and adolescence) (Robins, 1978; Wolfgang et al., 1972). Also, early onset of antisocial behaviour has been identified as a significant predictor of persistent antisocial behaviour (Tremblay et al., 1994).

Crucially, this subgroup of 'life-course persistent' offenders have poorer prognoses than do individuals beginning their criminal careers in adolescence or later life (Lahey et al., 1999; Loeber and Farrington, 2000; Moffitt, 1993a; Moffitt et al., 2002). Indeed, a negative correlation has been observed between age of onset of conduct problems and level of functional impairment (Lahey et al., 1999). Further, in adulthood these individuals continue offending and have erratic employment patterns in unskilled jobs, violent relationships with partners and few friends (Rutter et al., 1998).

1.1.1: Summary

Research has identified a subgroup of antisocial children who engage in crime from an early age and continue to present with chronic antisocial behaviour throughout adolescence and into adulthood. Early identification and neuro-cognitive characterisation of this subgroup would allow the development of targeted prevention efforts in early childhood. The next section will go on to describe psychopathy: a disorder that appears to successfully capture a sub-set of this group.

1.2: Introduction to psychopathy

In contrast with the diagnoses of conduct disorder (CD) and antisocial personality disorder (APD), psychopathy is not currently recognised in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (American Psychiatric Association; APA, 1994). DSM-IV antisocial behaviour disorders identify a heterogenous population. Importantly, only a subset of individuals presenting with CD and APD meet criteria for psychopathy.

1.2.1: The measurement of psychopathy; the PCL-R and APSD

The description of psychopathy originated with the work of Hervey Cleckley who delineated 16 diagnostic criteria (Cleckley, 1941). These included: superficial charm, lack of anxiety, lack of guilt, undependability, dishonesty, egocentricity, failure to form lasting intimate relationships, failure to learn from punishment, poverty of emotions, lack of insight into the impact of one's behaviour on others, and failure to plan ahead. Subsequently, Robert Hare developed the first formalized tool for the assessment of psychopathy in adults, which is now in its third edition (Psychopathy Checklist-Revised; PCL-R; Hare, 1991; Hare, 2003). Following observations of behavioural and neurocognitive consistencies between children and adults with the disorder, tools for the assessment of psychopathy in childhood and adolescence have also been developed. These include the Antisocial Process Screening Device (APSD; Frick and Hare, 2001) and the Psychopathy Checklist: Youth Version (Forth et al., in press; Kosson et al., 2002).[1] Whilst the PCL-R and APSD index a similar syndrome in adults and children (Frick et al., 1994; Harpur et al., 1989), involving affective-interpersonal and behavioural components, there exist some content differences between the measures. These differences, at least partially, reflect the intention of both tools to be age appropriate. Thus some PCL-R items have no

[1] In the present thesis psychopathic tendencies in children have been identified through use of the APSD so discussion will focus largely on this tool.

APSD counterparts (*e.g.* parasitic lifestyle), and some APSD items have no PCL-R counterparts (*e.g.* concerned about school work).

Both the PCL-R and the APSD consist of 20 items which assess affect and behaviour. The PCL-R is scored on the basis of an extensive file review and a semi-structured interview (administered by trained interviewers). The APSD is scored on the basis of parental and/or teacher ratings. For both the PCL-R and APSD each behavioural item is scored between 0 and 2 points, leading to a maximum possible score of 40. Adults scoring 30 or more on the PCL-R are generally considered psychopathic while those scoring less than 20 are considered non-psychopathic. There are less well established inclusion criteria for children, however cut-offs varying from 25-30 points for the psychopathic tendencies group and 10-20 points for the comparison group have typically been used (Blair et al., 2005; Blair et al., 2001a; Frick, 1995; Frick et al., 2000; Frick and Ellis, 1999; Frick and Hare, 2001).

Early factor analyses of the PCL-R and APSD derived similar solutions, comprising two inter-correlated factors (Frick, 1995; Frick et al., 2000; Frick et al., 1994; Hare, 1991; Harpur et al., 1988; Harpur et al., 1989). In adult studies (using the PCL-R) these were termed, factor 1; 'interpersonal/ affective', and factor 2; 'impulsive/ antisocial lifestyle' (Hare, 1991; Harpur et al., 1988; Harpur et al., 1989). In studies with children (using the APSD) they were termed, factor 1; 'callous and unemotional interpersonal style' (or *CU*), and factor 2; 'impulsivity/ conduct problems' (or *I/CP*) (Frick, 1995; Frick et al., 2000). The factors identified using the PCL-R and APSD are similar, with the factor 1 items referring to an emotional dysfunction (*i.e. CU* traits) and the factor 2 items referring to an impulsive and antisocial lifestyle. More recently 3-factor solutions have been offered in studies with adults and children (Cooke and Michie, 2001; Frick et al., 2000; Frick and Hare, 2001). Essentially, these have divided factor 1 into two components: an interpersonal component and an abnormal affect component (see tables *1.1.* & *1.2.*), whilst factor 2 has remained the same.

Researchers, using these tools, have identified a relatively small subset of antisocial individuals as presenting with psychopathy. Indeed, in contrast to high

prevalence rates reported for DSM-IV antisocial behaviour disorders, the diagnosis of psychopathy comprises only a small subset of antisocial individuals (Frick, 2000). For example while up to 80% of US inmates reach DSM-IV diagnostic criteria for APD, only 15 to 25% of US inmates meet criteria for psychopathy according to the PCL-R criteria (Fazel and Danesh, 2002; Hart and Hare, 1996). Preliminary epidemiological work with childhood community samples using the APSD has indicated a prevalence rate for psychopathic tendencies of between 1 and 3.5% (*i.e.* approximately one quarter of the community incidence rate of CD) (Frick, personal communication).

1.2.2: Criminality and Violence in Psychopathic Individuals

Importantly, the PCL-R and APSD have successfully identified a uniquely violent and criminal target group. Indeed, Kosson and colleagues (1990) demonstrated that the psychopathic offender commits more types of crimes, as well as more crimes of any type, relative to non-psychopathic offenders. Related to this, upon release from incarceration psychopaths are prone to recidivate (Grann et al., 1999; Hare et al., 2000; Hart et al., 1988; Hemphill et al., 1998; Serin and Amos, 1995). Hemphill, Wong & Hare (1998) observed that the correlation between psychopathy and recidivism was significantly higher than that between DSM-IV APD and recidivism. Furthermore, studies have often shown that psychopaths are more likely than non-psychopathic criminals to recidivate with a violent offence (Hare et al., 2000; Hare and McPherson, 1984). Hemphill, et al., (1998) examined 9 prospective studies of psychopathy and recidivism. Results showed that within a year of release, individuals with psychopathy were three times more likely to recidivate, and four times more likely to recidivate violently than non-psychopathic criminals.

In addition to a general increase in violent acts, it appears also that the *type* of violence associated with psychopathy may be different to that which is associated with other antisocial behaviour disorders. Specifically, psychopathy has been associated with high levels of predatory, instrumental violence *and* reactive violence (Cornell et al., 1996; Williamson et al., 1987; Woodworth and

Porter, 2002). A distinction between reactive and instrumental aggression has been made for some time in the scientific literature (Barratt et al., 1999; Barratt et al., 1997b; Berkowitz, 1993; Crick and Dodge, 1996; Dodge and Coie, 1987; Linnoila et al., 1983; Vitiello and Stoff, 1997). Whilst reactive aggression is initiated without regard for any potential goal, instrumental aggression (also referred to as proactive aggression), in contrast, is purposeful and goal directed. Goals, for example, may be to gain a victim's possessions or to increase status within a hierarchy) Woodworth and Porter (2002) demonstrated that psychopathic murderers were more likely to have committed a premeditated murder than non-psychopathic murders, who in contrast, were more likely to have committed a reactively aggressive, 'crime of passion', murder. The accumulating evidence linking high rates of predatory, violent, tendencies with psychopathy has been considered to reflect a lack of empathy in this population. Further, it has led to suggestions that the factor 1, affective component forms the core of the disorder (Barry et al., 2000; Blair, 2003a; Christian et al., 1997; Frick et al., 2000; Frick et al., 1994; Hart and Hare, 1997; Hawes and Dadds, in press; Viding et al., 2005). Indeed, the reason why most individuals with DSM-IV antisocial behaviour diagnoses do not fulfil the criteria for psychopathy is due to the absence of *CU* traits (Barry et al., 2000; Hart and Hare, 1997). Further, strong evidence for a substantial genetic component to *CU* traits has recently been reported (Viding et al., 2005).

table 1.1: Three-factor structure of the PCL-R

Factor 1: Deficient affective experience items	Factor 2: Arrogant and deceitful interpersonal items	Factor 3: Impulsive and irresponsible items	Items failing to load on any factor
6. Lack of remorse or guilt 7. Shallow affect 8. Callous/lacks empathy 16. Failure to accept responsibility for own actions	1. Glibness/Superficial Charm 2. Grandiose sense of self-worth 4. Pathological lying 5. Conning/Manipulative	3. Need for stimulation/proneness to boredom 9. Parasitic lifestyle 13. Lack of realistic, long-term goals 14. Impulsivity 15. Irresponsibility	10. Poor behavioural controls 11. Promiscuous sexual behaviour 12. Early behavioural problems 17. Many short-term marital relationships 19. Revocation of conditional release 20. Criminal versatility

table 1.2: Three-factor structure of the APSD

Factor 1: Callous and unemotional items	Factor 2: Narcissism items	Factor 3: Impulsivity items	Items failing to load on any factor/excluded from analysis
3. Concerned about schoolwork† 7. Keeps promises† 12. Feels bad or guilty† 18. Concerned about the feelings of others† 19. Does not show emotions 20. Keeps the same friends†	5. Emotions seem shallow 8. Brags excessively 10. Uses or cons others 11. Teases others 14. Can be charming, but seems insincere 15. Becomes angry when corrected 16. Thinks he/she is better than others	1. Blames others for mistakes 4. Acts without thinking 9. Gets bored easily 13. Engages in risky activities 17. Does not plan ahead	2. Engages in illegal activities 6. Lies easily and skilfully

†Items are reverse-scored

1.2.3: Co-Morbidity Between Psychopathic Tendencies and ADHD

Attention-deficit/hyperactivity disorder (ADHD) is another DSM-IV disorder that predicts the display of antisocial behaviour in childhood, adolescence and adulthood (Babinski et al., 1999; Farrington, 1990; Mannuzza et al., 1989; Simonoff et al., 2004). ADHD is defined as *"a persistent pattern of inattention and/or hyperactivity-impulsivity that is more frequent and severe than is typically observed in individuals at a comparable level of development"* (APA, 1994). Two independent components within ADHD are (i) inattention and (ii) hyperactivity-impulsivity. ADHD is one of the most common chronic disorders of childhood with rates in the US varying between 1% and 20% (DuPaul, 1991).

Recent reports have suggested that the diagnoses of psychopathic tendencies and ADHD are highly co-morbid (Barry et al., 2000; Colledge and Blair, 2001; Johansson et al., 2005; Lynam, 1996). For example Colledge and Blair (2001) observed that 75% of their sample of children with psychopathic tendencies also met criteria for ADHD. Interestingly, inter-correlations between ratings of psychopathic tendencies and ADHD were found to be due to an association between APSD factor 2 scores (*i.e.* impulsive and antisocial lifestyle items) and ADHD-defined impulsivity (Colledge and Blair, 2001). Indeed, hyperactivity-impulsivity (rather than inattention) in children with ADHD has been linked with later criminality and antisocial behaviour (Babinski et al., 1999; Barkley, 2002; Farrington and West, 1993). It is important to note that the combination of hyperactivity-impulsivity and antisocial behaviour should not be equated with psychopathy. Indeed hyperactivity-impulsivity and antisocial behaviour, without concomitant *CU* traits, are insufficient for a diagnosis of psychopathy (Barry et al., 2000; Hart and Hare, 1997; Viding et al., 2005). Due to reports of high co-morbidity between psychopathic tendencies and ADHD it is becoming increasingly important for researchers to be able to differentiate neuro-cognitive impairments associated with psychopathic tendencies and those associated with ADHD (Colledge and Blair, 2001).

1.2.4: Summary

Psychopathy is a developmental disorder that incorporates only a small subset of antisocial individuals. It has been associated with an increased risk for violence and criminality from childhood to adulthood. Factor analyses have indicated that the disorder of psychopathy is comprised of an abnormal affect (*CU*) component, which, it has been suggested, may form the core of the disorder, and an impulsive-antisocial behavioural (*I/CP*) component. It is noteworthy that children with psychopathic tendencies are likely to present co-morbidly with ADHD. The next section will go on to present and evaluate six theories attempting to explain psychopathy, along with empirical research and observations that prompted them.

1.3: Theories of Psychopathy

This section will present six theories attempting to explain psychopathy. It will also assess their ability to account for the current neuro-cognitive data and behavioural observations associated with this population. Firstly it will discuss two theories focusing upon the emotional dysfunction observed in individuals with psychopathy. Secondly it will discuss two theories emphasizing the involvement of prefrontal brain regions in the expression of psychopathy. Thirdly it will present an attention-based account, and finally, it will present a theory attempting to integrate the emotion dysfunction and fear dysfunction positions.

1.3.1: The Violence Inhibition Mechanism Model

The violence inhibition mechanism (VIM) model of psychopathy (Blair, 1995; Blair et al., 1997) emphasizes the importance of empathy for the normal development of morality. It was devised in light of data demonstrating that many social animals, including humans, appear to find the distress of conspecifics aversive (Church, 1959; Masserman et al., 1964; Rice, 1965; Rice and Gainer,

1962). As regards humans, the model suggests that the experience of conspecific fear or sadness is an innate aversive stimulus. Furthermore it is suggested that this innate mechanism reduces the probability of an individual engaging in actions that previously led to the distress of another person. It is suggested that, when activated by distress cues, the VIM results in increased autonomic activity, attention and activation of the brainstem threat response system (Blair 1995). According to the model, moral socialization occurs when activation of the mechanism is paired with representations of the acts that caused the distress (Blair, 1995). Accordingly, by means of association, internal representations of moral transgressions also become triggers for the VIM. The normally developing child therefore initially finds the distress of other individuals aversive. Through socialization thoughts of acts that may cause distress to others also become aversive. It is proposed that this system is dysfunctional in individuals with psychopathy (Blair, 1995).

There is much evidence in support of the VIM theory as a model for human moral development. For example, the distress of others is considered aversive by most humans (Bandura and Rosenthal, 1966). Furthermore, the presentation of cues demonstrating another individual's fear or sadness reduces the probability of physical aggression (Perry and Perry, 1974). Moreover, healthy developing individuals, by the age of 3 years, show successful performance on the moral/conventional distinction test (Smetana and Braeges, 1990). In this task participants must decide whether transgressions described in vignettes are of a moral (usually victim-based) or conventional (social order-based) nature (Turiel, 1983). As predicted by the VIM position, individuals with psychopathy, even as adults, perform abnormally on the moral/conventional distinction test (Blair, 1995; Blair, 1997; Blair et al., 1995a; Blair et al., 2001d). While they do generally regard moral transgressions as more serious than conventional transgressions, they are less likely to make reference to the victims when explaining *why* this should be the case (Blair, 1995; Blair, 1997; Blair et al., 1995a; Blair et al., 2001d). In addition, when rules prohibiting the transgressions are removed, individuals with psychopathy are less likely to make a distinction

between moral and conventional transgressions (Blair, 1995; Blair et al., 2001d; Nucci and Herman, 1982). Further, in a task assessing emotion attributions, individuals with psychopathy have displayed anomalous concepts for guilt (but not for happiness, sadness or embarrassment) suggesting that their experience of this emotion may be abnormal (Blair et al., 1995b).

Additional support for the model comes from studies showing that appropriate empathic responses to victims lead to reduced levels of antisocial behaviour (Eisenberg et al., 1996; Feshbach, 1987; Perry and Perry, 1974). Indeed, one of the defining criteria of psychopathy, as indexed by both the PCL-R and APSD, is low empathy. Impairments in empathic responsiveness in individuals with psychopathy include reduced autonomic responses to the sadness of other individuals (Aniskiewicz, 1979; Blair, 1999; Blair et al., 1997; House and Milligan, 1976; Sutker, 1970) and impaired recognition of fearful and sad facial expressions and vocal affect (Blair et al., 2001c; Blair and Coles, 2000; Stevens et al., 2001). Notably, adult individuals with psychopathy and children with psychopathic tendencies do not present with impaired recognition of angry, happy, or surprised facial or vocal expressions (Blair et al., 2001c; Blair and Coles, 2000; Stevens et al., 2001).

In conclusion the VIM model provides a successful account of the development of morality and also of the emergence of instrumental antisocial behaviour in individuals with psychopathy. Furthermore, it successfully predicts impairment on certain tasks assessing morality and the processing of empathy cues. However it is unable to account for the range of impairments shown by individuals with psychopathy outside of the realm of empathy and moral development. Indeed, it is unable to explain much of the neuro-cognitive data associated with the fear dysfunction positions that will be presented in the next section. Further, the VIM model is under-specified at the neural level. As such, this model has essentially been superseded and re-formulated into the Integrated Emotion Systems model (see section *1.3.6.*).

1.3.2: The Fear Dysfunction Hypotheses

In an attempt to explain psychopathy, various researchers have made reference to the 'fear system'. These hypotheses have asserted that the emotional impairment observed in psychopathy is due to dysfunction in the neuro-physiological systems modulating fear-related behaviour (Cleckley, 1976; Eysenck, 1964; Fowles, 1988; Gray, 1987; Lykken, 1995; Mealey, 1995; Patrick, 1994; Pichot, 1978; Trasler, 1978; Trasler, 1973). David Lykken was one of the earliest theorists to associate psychopathy with reduced fearfulness (Lykken, 1957). Lykken contended that the psychopathic individual *"has an attenuated experience, not of all emotional states, but specifically anxiety or fear"* (Lykken, 1995, pg. 118). Essentially it was suggested that reduced fearfulness negatively impacts socialization and leads to the development of psychopathy. The explanation is as follows; in contrast to individuals with psychopathy, healthy individuals are frightened by punishment, and, during socialization, fear of punishment becomes associated with the action that resulted in punishment. This makes healthy individuals less likely to engage in that particular action in the future. It is hypothesized that individuals with psychopathy, because they are less aversively aroused by punishment, form weaker associations between a transgression and the consequent punishment, which in turn leads to an decreased propensity to avoid previously punished behaviours. Importantly, this process assumes that moral socialization is achieved through the use of punishment (Eysenck and Gudjonsson, 1989; Trasler, 1978).

The fear positions are able to account for a variety of empirical data from studies with psychopaths. Indeed, many of the earliest experimental investigations of psychopathy were prompted by the fear dysfunction hypotheses (e.g. Lykken, 1957). Specifically, these theories predict abnormally reduced emotional responses to aversive stimuli. Indeed, individuals with psychopathy have demonstrated reduced emotional responsiveness in a variety of experimental procedures such as, autonomic responses to aversive conditioning (Flor et al., 2002; Hare and Quinn, 1971; Lykken, 1957), anticipation of punishment (Hare, 1965; Hare, 1982; Hare et al., 1978; Ogloff and Wong, 1990), imagining

threatening events (Patrick et al., 1994) and augmentation of the startle reflex by aversive primes (Lang et al., 1990; Levenston et al., 2000; Pastor et al., 2003; Patrick, 1994).

Additionally the fear positions predict impaired performance by individuals with psychopathy in tasks of passive avoidance learning, reversal learning and extinction; all tasks in which punished responses signal that a behavioural action is inappropriate. Adult psychopaths have consistently presented with impairment in investigations of passive avoidance learning (Kosson et al., 1990; Newman and Kosson, 1986; Newman et al., 1990; Newman and Schmitt, 1998; Thornquist and Zuckerman, 1995). Their behaviour is characterised by an increased rate of approach toward stimuli predictive of punishment as compared with controls. The case as regards passive avoidance learning in children with psychopathic tendencies, however, is less clear as conflicting results have been reported (Newman et al., 1985; Scerbo et al., 1990). Also in line with the predictions of the fear position, adult individuals with psychopathy have presented with impairment in tasks assessing reversal learning and extinction such as the reversal component of the intra-dimensional/extra-dimensional (ID/ED) task, Bechara's Iowa gambling task and Newman's card extinction task (Mitchell et al., 2002; Newman et al., 1987). As with passive avoidance learning, however, the case as regards reversal learning ability in children with psychopathic tendencies is somewhat unclear. Specifically, while they have presented with impairment in some investigations, they have performed comparably to controls in others (Blair et al., 2001a; Fisher and Blair, 1998; O'Brien and Frick, 1996).

Despite considerable empirical success, the fear dysfunction hypotheses face several problems. Implicit in these hypotheses is the tenet that socialization is achieved through punishment. This assumption has been questioned (Blackburn, 1988; Blair and Morton, 1995). Instead the developmental literature indicates that moral socialization is achieved through the induction and fostering of empathy (Hoffman, 1984). Indeed, data discussed in section *1.3.1.* indicate that healthy developing children are able to distinguish between moral and

conventional transgressions at an early age (Smetana, 1981; Smetana, 1985; Smetana, 1993). The fear position is unable to explain this distinction, instead it follows that the only way to distinguish between 'good' and 'bad' would be according to whether the action had been previously punished or not. Further, conditioning theory (*e.g.* Dickinson, 1980) would predict that rather than the transgression it would be the *individual delivering the punishment* that would become associated with the punishment. Essentially, the punisher would become a highly predictive conditioned stimulus [CS] due to consistent temporal contiguity with the punishment. This is in contrast with the transgression which would not always be temporally contiguous with the punishment, and thus a poorly predictive CS. Indeed evidence suggests that children subject to corporal punishment do display fear toward the punisher rather than fear toward committing the transgression in question (Hoffman, 1994). Also the developmental literature indicates that moral socialisation is better achieved through the induction and fostering of empathy than through harsh authoritarian practices which rely on the use of punishment (Baumrind, 1971; Baumrind, 1983; Hoffman and Saltzstein, 1967). Indeed, there have been suggestions that while empathy facilitates moral socialisation, fear actually hinders it (Brody and Shaffer, 1982; Hoffman, 1994).

Additionally, many theories of fear dysfunction are under-specified at both the cognitive and neural levels, with few details offered concerning the computational properties of the system. The behavioural inhibition system (BIS) model (Gray, 1987; Gray and McNaughton, 1996; McNaughton and Gray, 2000) is one of the few detailed accounts of a fear system that has been used to explain psychopathy. In this account the BIS is thought to generate autonomic responses to punishing stimuli (through a process of classical conditioning) and also to inhibit responses following punishment (through a process of instrumental conditioning). Importantly, the BIS model assumes that there is a unitary fear system. The empirical literature, however, strongly suggests that instead of a single fear system there exist a series of at least partially separable neural systems that are engaged in specific forms of processing (which may be subsumed under

the term fear) (Amaral, 2001; Blair and Cipolotti, 2000; Killcross et al., 1997; Prather et al., 2001). For example, aversive conditioning and instrumental learning are two forms of processing in which the fear system is thought to be involved (Lykken, 1995; Patrick, 1994), yet the neural circuitry to achieve aversive conditioning and instrumental learning are doubly dissociable (Killcross et al., 1997). Whilst lesions to the central nucleus of the amygdala produce a deficit exclusive to aversive conditioning, lesions to the basolateral nucleus produce a deficit exclusive to instrumental learning (Killcross et al., 1997). Furthermore, early amygdala lesions lead to a massive reduction in neo-phobia but an increase in social phobia (Amaral, 2001; Prather et al., 2001).

In conclusion the fear positions have generated a considerable body of data, much of which is compatible with their hypotheses. However some of the assumptions upon which these positions rest have been questioned. In particular the assertion that moral development is achieved by means of punishment has been contested. Further, importantly, the assumption that there exists a unitary fear system appears to be false. Additionally, most theories of fear dysfunction are under-specified at both the cognitive and neural levels.

1.3.3: The Frontal Lobe Dysfunction Hypothesis

The term 'executive functions' refers to those cognitive processes that underlie flexible goal-directed behaviour, such as inhibiting dominant responses; and creating, maintaining and temporally sequencing behaviours (Burgess et al. 1998). Historically, investigations assessing executive functions have preferentially assessed those functions thought to rely, to a large degree, upon the integrity of the dorsolateral prefrontal cortex (DLPFC) such as tasks involving the planning, monitoring or inhibition of prepotent behaviours (Smith and Jonides 1999). While this may be a somewhat simplistic view, neuropsychological, functional imaging, and animal lesion evidence indicates that different aspects of executive functions may be dissociable and mediated by, at least partially, distinct neural systems subserved by different regions of the prefrontal cortex (Luria 1966;

Fuster 1980; Roberts, Robbins and Weiskrantz, 1998; Shallice 1988; Baddeley and Della Sala 1998). Thus evidence suggests that certain functions may rely *more heavily* upon certain regions over others within the prefrontal cortex (PFC).

Abnormal executive functioning, as a result of frontal lobe dysfunction, has been associated with antisocial behaviour (Barratt, 1994; Elliot, 1978; Gorenstein, 1982; Moffitt, 1993a; Raine, 1997; Raine, 2002a). Consequently, this association has led to suggestions that frontal lobe dysfunction may be the cause of psychopathy in particular, but also antisocial behaviour more generally (Gorenstein, 1982; Moffitt, 1993a; Raine, 2002a; Raine, 2002b; Toupin et al., 2000). Indeed reviews of the experimental literature have often concluded that executive functioning is impaired in adults and children with antisocial behaviour (Dolan and Park, 2002; Kandel and Freed, 1989; Moffitt, 1993b; Morgan and Lilienfield, 2000; Pennington and Ozonoff, 1996).

In contrast to other antisocial groups, however, individuals with psychopathy have not shown executive dysfunction on measures linked primarily to DLPFC functioning such as the Wisconsin Card Sorting Task (LaPierre et al., 1995), the Controlled Oral Word Association Test (Roussy and Toupin, 2000; Smith et al., 1992) and the higher-order shift stages of the Intra-dimensional/Extra-dimensional (ID/ED) Shift task (Mitchell et al., 2002). Instead there are indications that individuals with psychopathy are impaired on measures of frontal functioning that have been preferentially linked to the integrity of orbital and ventrolateral PFC, such as tasks of reversal learning and extinction (introduced in the section *1.3.2.*) (Mitchell et al., 2002; Newman et al., 1997; Roussy and Toupin, 2000). Indeed, as will be discussed further, patients sustaining lesions to these areas also present with reversal learning impairments (Bechara et al., 1994; Bechara et al., 1999; Bechara et al., 1997; Bechara et al., 2001; Berlin et al., 2004; Fellows and Farah, 2003; Hornak et al., 2004; Rolls et al., 1994).

In addition to the neurocognitive similarities described above, individuals with psychopathy appear to share a range of social behaviours with patients who have sustained damage to orbital, medial or ventral PFC. Patients with acquired

lesions of these regions, for example, often present with emotional and personality changes such as irresponsibility, lack of concern for the present or future and increased aggression (Anderson et al., 1999; Barrash et al., 2000; Blair and Cipolotti, 2000; Berlin et al., 2004; Grafman et al., 1996; Hecaen and Albert, 1978; Stuss and Benson, 1986). Indeed some of these behaviours are consistent with the diagnosis of psychopathy (see tables *1.1.* and *1.2.*). It must be noted, however, that the type of aggression to which patients with lesions to the ventral PFC are prone is exclusively reactive in nature (Anderson et al., 1999; Damasio, 1994; Grafman et al., 1996; Pennington and Bennetto, 1993). As noted in section *1.2.2.*, reactive aggression is usually triggered by a frustrating or threatening event and is not goal-directed (Dodge and Coie, 1987; Panksepp, 1989; Panksepp, 1998). In contrast, individuals with psychopathy are renowned for displaying high rates of reactive *and* proactive aggression (Blair, 2001; Blair, 2003a; Cornell et al., 1996; Williamson et al., 1987; Woodworth and Porter, 2002). Human investigations have indicated that these two forms of aggression are mediated by at least partially separable neural systems (Barratt et al., 1999; Barratt et al., 1997a; Berkowitz, 1993; Linnoila et al., 1983). In the case of the reactive aggression occasionally displayed by ventral-frontal patients, dysfunction within these areas might lead to dysregulated modulation of the brainstem systems that mediate the basic response to threat, in turn increasing the probability of reactive aggression (Blair, 2004). In keeping with this, a positron emission tomography study of patients with personality disorders revealed reduced regional cerebral blood flow in orbitofrontal cortex (OFC) which negatively correlated with a history of (mostly reactive) aggression (Goyer et al., 1994).

At the theoretical level, the frontal lobe dysfunction hypothesis lacks a sound argument of why lesions to the frontal lobes would increase the probability of psychopathy. Most accounts do not adequately specify a mechanism by which frontal cortex dysfunction might lead to this disorder. Frequently, reference is made to 'reduced inhibition' and dysfunction in 'inhibitory mechanisms' following frontal dysfunction. However reduced inhibition is unable to account

for the full neurocognitive profile and behavioural sequelae associated with psychopathy.

In conclusion, the frontal lobe positions attempt to describe the association between frontal damage and psychopathy or more generally, antisocial behaviour. While these positions are successful in linking antisocial behaviour with executive dysfunction they require further specification in order to describe the link between frontal dysfunction and aggression. Specifically, aggression has been linked with damage to orbital and ventral regions, and not dorsolateral regions. Thus, with refinement, these positions would successfully describe the association between OFC/ ventrolateral PFC damage and reactive aggression. The frontal lobe positions do not attempt to account for other behavioural and neurocognitive observations associated with psychopathy. Most notably, the instrumental violence for which psychopathy is renowned is unaccounted for by these positions.

1.3.4: The Somatic Marker Hypothesis

The somatic marker (SM) hypothesis is an account of the functions of the ventromedial frontal cortex (Bechara et al., 2000a; Damasio, 1994). According to this position, the ventromedial frontal cortex acts as a repository, and is involved in the formation of linkages between factual knowledge and bio-regulatory states (Bechara et al., 2000a; Damasio, 1994). When emotionally significant decisions are being made (for example decisions involving rewards or punishments), bio-regulatory states provide 'affective colouring' that automatically biases the individual's response. In short, bodily feedback (*i.e.* somatic markers) rapidly labels particular options as either good or bad, thereby influencing the likelihood that a particular response will be made. This labelling can occur via a 'body loop' whereby a SM is conveyed to somatosensory cortices, but it can also occur via an 'as-if body loop' in which the body is bypassed and reactivation signals are conveyed to the somatosensory structures. In short, the somatosensory pattern marks the scenario as either good or bad, allowing the rapid rejection or

endorsement of specific option-outcome pairs. It has been suggested that a dysfunctional SM system produces a syndrome known as 'acquired sociopathy'. Further, Damasio has suggested that psychopathy might be the developmental form of acquired sociopathy (Damasio, 1994; Damasio et al., 1990).

The SM position was developed following observations that patients with lesions to the ventromedial frontal cortex fail to show autonomic responses to visually presented social stimuli, such as scenes of social disaster, mutilation and nudity (Damasio et al., 1990; Damasio et al., 1991). Also patients with lesions to ventromedial frontal cortex have performed poorly on Bechara's Iowa gambling task, which was designed as a test of this position (Bechara et al., 1994; Bechara et al., 1999; Bechara et al., 1997; Bechara et al., 2001; Bechara et al., 2000b). Essentially, these patients, in contrast to controls, approach risky decks at high rates and do not produce anticipatory skin conductance responses (SCRs) when doing so. They do, however, produce outcome-related SCRs. These data have been interpreted as demonstrating a lack of somatic markers in this group.

As indicated in section *1.3.2.*, in line with the suggestion that psychopathy is a developmental form of acquired sociopathy, individuals with psychopathy present with impaired performance on the Iowa gambling task (Blair et al., 2001a; Mitchell et al., 2002). However, whilst the SCRs of psychopaths have not been measured during performance of the Iowa gambling task, unlike ventromedial patients individuals with psychopathy do not present with generally reduced autonomic responses to visually presented social stimuli (Blair, 1999; Blair et al., 1997; Levenston et al., 2000; Patrick et al., 1993). It thus appears that individuals with psychopathy *do* generate somatic markers even if they do not use them effectively on tasks such as the Iowa gambling task. Also, as discussed in section *1.3.3.*, lesions of the ventral PFC lead to elevated levels of reactive (and not instrumental) aggression (Anderson et al., 1999; Barrash et al., 2000; Blair and Cipolotti, 2000; Burgess and Wood, 1990; Grafman et al., 1996; Pennington and Bennetto, 1993). As such, any account linking ventromedial function with antisocial behaviour would have to explain this dissociation. As a unitary account of antisocial behaviour the somatic marker hypothesis is unable to explain the

increased preponderance of reactive aggression in patients with damage to the somatic marker system.

In conclusion, the SM hypothesis is an interesting model of ventromedial PFC functioning. However, its application to the understanding of aggression and antisocial behaviour has been less successful. Predictions of the functional deficits of individuals with psychopathy have only been partially confirmed.

1.3.5: The Response-Set Modulation Hypothesis

The response-set modulation (RM) hypothesis is an attentional account of psychopathy (Newman, 1998; Patterson and Newman, 1993). Newman and colleagues suggest that response-set modulation involves *"a rapid and relatively automatic (i.e., non-effortful or involuntary) shift of attention from the effortful organization and implementation of goal-directed behaviour to its evaluation"* (Newman et al., 1997, pg. 564). Further, it is suggested that this *"brief and highly automatic shift of attention ... enables individuals to monitor and, if relevant, use information that is peripheral to their dominant response set (i.e., deliberate focus of attention)"* (Lorenz and Newman, 2002, pg. 92). According to the model individuals with psychopathy are predisposed to function at a 'non-effortful' level of self-regulation. When functioning at this level, however, it is suggested that they fail to process peripheral information that would serve to improve task performance. Dysfunction within the system responsible for response-set modulation will result in impoverished performance under conditions where a salient stimulus ought to divert attention from on-going behaviour. Gorenstein and Newman (Gorenstein and Newman, 1980) proposed a physiological animal model of psychopathy that is based around Gray's 'Septo-hippocampal formation' system (Gray, 1972). This circuit includes the medial septum, posterior hippocampus, and OFC (Gorenstein and Newman, 1980). In the animal literature, RM deficits were characterised as a failure to inhibit approach responses despite punishment, extinction or contingency reversal.

The RM hypothesis has generated a considerable body of experimental work. For example, Newman and colleagues introduced the passive avoidance paradigm and the card extinction task, both of which individuals with psychopathy fail (Blair et al., 2004; Fisher and Blair, 1998; Newman et al., 1987; O'Brien and Frick, 1996). According to the RM hypothesis, the poor performance of individuals with psychopathy on these tasks relates to their inability to shift their attention, from their goal of responding to gain reward, to the peripheral punishment information (Lorenz and Newman, 2002; Newman, 1998; Newman et al., 1990; Newman et al., 1987; Patterson and Newman, 1993). A similar explanation would be used to describe the poor performance of individuals with psychopathy on reversal learning tasks.

The RM hypothesis has been associated with the development of an assortment of interesting paradigms. However it faces difficulties. In particular, while the RM hypothesis is an attentional account, it is unclear to what extent this account is compatible with contemporary models of attention. For example, in tasks of passive avoidance learning, stimuli are presented serially and feedback information is presented, *independently of any other information*, that is, following responses the feedback is presented in the absence of potentially distracting information. According to attentional models (*e.g.* Desimone and Duncan, 1995; Lavie, 1995) it would be difficult to account for a lack of attention to this information (given the absence of competing stimuli). As such, the fact that the punishment information does not appear to modulate the behaviour of individuals with psychopathy would seem to suggest that these individuals have difficulties *learning from punishing information*, rather than that they are unable to attend to this information.

In conclusion, the RM hypothesis has resulted in the development of an assortment of interesting paradigms. However, it is unclear the extent to which this attention-driven hypothesis is compatible with contemporary positions on attention. Further, this position is unable to account for data relating to empathy and morality (see section *1.3.1.*).

1.3.6: The Integrated Emotion Systems Model

The integrated emotion systems (IES) model (Blair, 2004), a neuro-cognitive theory, may be considered an extension of the VIM and fear positions (Blair, 2003a; Blair, 2003b; Blair, 2004). This theory suggests that the primary dysfunction, with regard to psychopathy, is within the amygdala but that dysfunction also exists within OFC/ventral PFC.

This model assumes a fundamental impairment in the representation of affect that is implemented by the amygdala. These affect representations are also thought to be required for the successful processing of fearful and sad expressions (Aniskiewicz, 1979; Baird et al., 1999; Blair et al., 2001c; Blair et al., 2002; Blair et al., 1997; Blair et al., 1999; Breiter et al., 1996; Drevets et al., 2000; Morris et al., 1996; Phillips et al., 1998; Phillips et al., 1997; Schneider et al., 1994) and the appropriate development of moral socialization (Blair, 1995; Blair et al., 1997) (*i.e.* functions explained by the VIM account, see section *1.3.1.*). Further, the amygdala is also considered to be required for successful acquisition of Pavlovian-type associations such as aversive conditioning and startle-reflex modulation (*i.e.* functions explained by the fear accounts, see section *1.3.2.*).

This amygdala-based Pavlovian-type learning is represented within the IES model by two modules of non-linear, computational units, with one module representing the amygdala and one module representing sensory regions (*e.g.*, auditory, visual and temporal cortex) (see figure *1.1.*). The connections between the units in the different modules are reciprocal, reflecting the bi-directional interconnections of the amygdala with cortical regions (Amaral et al., 1992). The strength of the connections between units in the different modules increase through Hebbian Learning (Hebb, 1949). Indeed, recent data at the cellular level provides evidence that learning within the amygdala may indeed be of a Hebbian nature (Blair et al., 2001b).

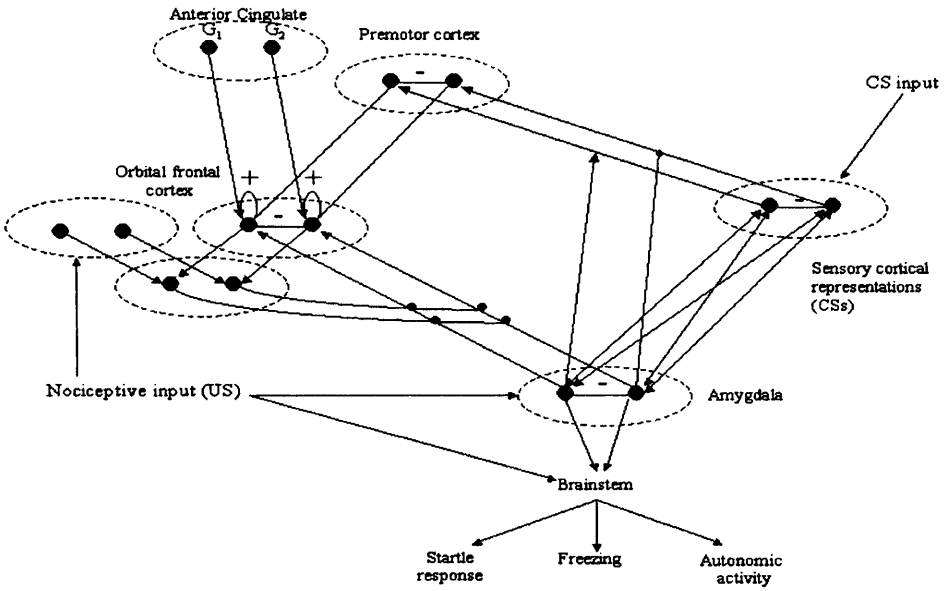
In terms of instrumental learning, the IES model makes a critical distinction between emotional learning that is hypothesized to be amygdala-reliant and that which is hypothesized to be amygdala-independent (Blair, 2004). In

stimulus-reinforcement learning (*i.e.* the formation of stimulus-reward and stimulus-punishment associations; see section *2.1.*), an association must be made between the stimulus and a valence-representation of the outcome, that is, its intrinsic motivational value (Baxter and Murray, 2002). This type of learning is thought to be amygdala reliant (Baxter and Murray, 2002). Animal work has indicated that stimulus-reinforcement learning requires interaction between a neural circuit including amygdala and medial OFC (Ambrogi Lorenzini et al., 1991; Bermudez-Rattoni et al., 1997; Bermudez-Rattoni and McGaugh, 1991; Cahill and McGaugh, 1990; Everitt et al., 2003; Schoenbaum et al., 2003; Treit and Menard, 1997). In contrast, in the formation of stimulus-response associations, an association is simply made between a stimulus and motor response (Baxter and Murray, 2002). This type of learning is thought to be amygdala-independent (Baxter and Murray, 2002). Instead, animal investigations have suggested that it is reliant upon a circuit including temporal cortex and caudate (Messinger et al., 2001; Packard, 1999; Packard and McGaugh, 1996).

Within the IES model, instrumental learning is represented using three modules of computational units. The first (non-amygdala reliant) module corresponds to units coding motor responses and includes premotor cortex and basal ganglia. The second (amygdala-reliant) module corresponds to units coding expectation of reward or punishment and represents medial OFC. It is suggested that units in medial OFC receive information in order to solve response competition on the basis of not only the activation of premotor units but also expectations of reinforcement that are provided by the amygdala. In addition, they receive input from units from the third module, which involves anterior cingulate, and represents desired goal states. In support of this formulation, empirical data suggests that amygdala lesions do impair instrumental learning (see above) (Ambrogi Lorenzini et al., 1999; Everitt et al., 2000; Killcross et al., 1997; LeDoux, 2000; Treit and Menard, 1997). Moreover, individuals with psychopathy have shown impairment on measures of instrumental learning (Fine et al., submitted) and passive avoidance learning (Newman and Kosson, 1986; Newman and Schmitt, 1998; Thornquist and Zuckerman, 1995). Interestingly,

recent data indicates that the hypothesized stimulus-punishment impairment is more pronounced than the stimulus-reward-impairment (Peschardt et al., submitted).

figure 1.1: The IES model (Blair, 2004)



Sensory cortex (auditory, visual and temporal cortex) and the hippocampus allow the representation of conditioned stimuli. Contiguous activation of representations of conditioned stimuli in sensory cortex and amygdala activation by an unconditioned stimulus will increase the connections between the two representations through Hebbian learning, allowing the CS to activate the brainstem even if the US is not present. Expectations of reinforcement transmitted from the amygdala to medial OFC allow resolution if more than one motor response option has been activated. Goal representations also modulate this processing. It is suggested that there are comparator units in ventrolateral PFC that would detect mismatches between expectations of reinforcement (provided by the amygdala units) and actual reinforcement. When activated these would disrupt the connections (weights) between amygdala units and OFC units as a function of the degree of the previous strength of these connection weights.

The IES model also incorporates comparator units in ventrolateral PFC which detect mismatches between expected and actual reinforcement and allow successful performance on tasks of reversal learning and extinction. When activated, these units disrupt the connection weights between amygdala and medial OFC units. Importantly, this is expected to occur as a function of the degree of the previous strength of these connection weights. Thus, under conditions where reinforcement had been a certainty and the connection weights were high, there would be considerable disruption. Under conditions where the reinforcement contingency was less obvious and the connection weights were lower, there would be less disruption. This process allows another unit to develop the new expectation of reinforcement, associated with the new contingency. The role of ventrolateral PFC in reversal learning, following prediction error theory (O'Doherty et al., 2004; Schultz et al., 1997; Sutton and Barto, 1981), is viewed as a function of the degree to which there is a mismatch between the expectation of reinforcement and the occurrence of reinforcement. Moreover, the greater the degree of dysfunction, the more difficult it would be for the individual to identify the contingency change. Thus, ventrolateral PFC is thought to be particularly involved in the detection of contingency change and the gating of motor responding with reference to that contingency change.

In conclusion, the IES model successfully accounts for much data, including those data derived from other positions, such as the VIM, fear, SM, and RM positions. Particularly important for this thesis – the IES position provides an alternative account for the reversal learning and passive avoidance deficits observed in psychopathic individuals.

1.3.7: Summary

Section *1.3.* presented six theories attempting to explain psychopathy. While the violence inhibition mechanism (VIM) theory provided a successful account of the development of morality, and the adverse consequences of an impaired system, it was unable to account for the emotional learning data derived

from other positions. Conversely, the fear positions provided an explanation for much of the emotional learning data, but were unable to explain data derived from the VIM position. Further, the fear positions rest upon the questionable tenet that moral socialisation is achieved through punishment. Next, the frontal lobe dysfunction positions provided an under-specified account of the association between frontal dysfunction and aggression, although they were able to describe the association between psychopathy and *reactive* aggression. The somatic marker (SM) position provided an unsuccessful account of psychopathy – in particular, it appears that psychopaths are able to generate 'somatic markers'. Further this position was unable to explain the preponderance of instrumental aggression displayed by psychopaths. The response-set modulation (RM) hypothesis introduced an assortment of experimental paradigms, including the passive avoidance learning paradigm, however this attentional account of psychopathy was incompatible with current models of attention. Finally the IES model attempted to integrate the VIM and fear positions. This position appeared to successfully account for a wide range of data, including those prompted by the other positions.

1.4: Summary and Aims

In summary Chapter 1 introduced the phenomena of antisocial behaviour and psychopathy, which is a developmental disorder characterized by display of violent, aggressive and criminal behaviours. Next theories attempting to explain psychopathy were discussed and evaluated on the basis of empirical data obtained with this population and theoretical issues. The remainder of this thesis will be devoted to the investigation of two neurocognitive tasks where conflicting data has been reported between children with psychopathic tendencies and adult psychopaths, and also where there exist alternative explanations of the dysfunction: passive avoidance learning and reversal learning.

Chapter 2 - Passive Avoidance Learning in Children with Psychopathic Tendencies

2.1: Introduction to Passive Avoidance Learning

Passive avoidance learning is the ability to avoid stimuli predictive of punishment. Human studies of passive avoidance learning involve randomized, serial presentation of stimuli on a computer screen. Participants are required to avoid conditioned stimuli predictive of punishment (CS-s) and approach conditioned stimuli predictive of reward (CS+s). On each trial participants must decide whether to respond to the CS (responses are usually made by button-press). Consequently they receive reward or punishment (contingent upon correct or incorrect approach behaviour respectively). Importantly participants are not informed of reward- and punishment-contingencies prior to the experimental procedure. Instead, they must learn by trial-and-error responding.

figure 2.1: A depiction of all possible outcomes in a standard task of passive avoidance learning

		Action	
		Approach	*Avoid*
Accuracy	*Correct*	Hit	CR
	Incorrect	PAE	OE

Correct responses are highlighted in green and incorrect responses are highlighted in red. Key: CR = correct rejection, PAE = passive avoidance error, OE = omission error.

It has been recently hypothesized that the most effective method of performing this type of instrumental learning involves the formation of stimulus-reinforcement associations (Baxter and Murray, 2002) (see section *1.3.6.*). In explanation, the stimuli appear serially, and are consistently associated with either

reward or punishment, this is thought to lead to the development of an association between the sensory features of the stimulus and the related affective value (*i.e.* the feedback information) (Blair, 2004). In short, a stimulus is 'tagged' as either 'good' or 'bad'. This conceptualisation of affect-driven learning may be contrasted with, motor-driven, stimulus-response learning (see sections *1.3.6.* and *4.1.)*.

The experimental design of passive avoidance learning tasks is typically a 2 (Action) x 2 (Accuracy) factorial (see figure *2.1.*). Thus individuals are able to make two types of correct response: correct approaches and avoidances (termed hits and correct rejections respectively). Likewise, they may make two types of incorrect response: incorrect approaches and avoidances (termed passive avoidance errors and omission errors) respectively.

2.2: Experiment 1

As introduced in the previous chapter, adult individuals with psychopathy have consistently demonstrated impaired passive avoidance learning (Newman and Kosson, 1986; Newman et al., 1990; Newman and Schmitt, 1998; Thornquist and Zuckerman, 1995). The case is less clear, however, as regards children with psychopathic tendencies (Newman et al., 1985; Scerbo et al., 1990). Experiment 1 aims to investigate and further characterise the passive avoidance learning difficulties presented by children with psychopathic tendencies.

The computerized number passive avoidance task, introduced by Newman and Kosson (Newman and Kosson, 1986), is the most frequently used measure of passive avoidance learning with psychopathic individuals. In this task, participants are presented with a series of two-digit numbers some of which, when approached, result in reward while others result in punishment (Kosson et al., 1990; Newman and Kosson, 1986; Newman et al., 1990). In the original investigation using this task adult psychopaths were found to commit more passive avoidance errors than comparison individuals (Newman and Kosson, 1986). This finding has been consistently replicated (Kosson et al., 1990; Newman et al., 1990; Thornquist and Zuckerman, 1995). Similarly, one study, using a non-computerized version of the passive avoidance learning task, reported impairment in passive avoidance learning in adolescents with psychopathic tendencies (Newman et al., 1985). In contrast, however, a second study reported no significant group differences in rates of passive avoidance errors; instead, the adolescents with psychopathic tendencies made significantly fewer misses than the comparison group (Scerbo et al., 1990).

Importantly, neither of the two existing investigations assessing passive avoidance learning ability of children with psychopathic tendencies assessed the potential influence of attention-deficit/hyperactivity disorder (ADHD). Given recent reports of high co-morbidity between these groups (Colledge and Blair, 2001) it is becoming increasingly important for researchers to exercise experimental controls in order to differentiate neuro-cognitive impairments

associated with psychopathic tendencies and those associated with ADHD. Indeed, passive avoidance learning has also been assessed in children with ADHD, with these studies also providing conflicting results (Hartung et al., 2002; Iaboni et al., 1995; Milich et al., 1994). Milich et al (1994) found, for boys only, that ADHD symptoms were significantly positively correlated with passive avoidance error rates. Iaboni et al (Iaboni et al., 1995) also found that individuals with ADHD presented with increased passive avoidance errors than comparison individuals. More recently, Hartung et al (Hartung et al., 2002) found that the observed association between ADHD and passive avoidance learning impairments was in fact due to the association between ADHD and conduct disorder (CD).

2.2.1: Task Manipulation

As described in section *2.1.*, it has been suggested that passive avoidance learning requires participants to form appropriate stimulus-reinforcement associations. Following the IES model (Blair, 2004), a potential mechanism for this is Hebbian Learning (Hebb, 1949).

``When the axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased''.

The Hebbian Learning Rule (Hebb, 1949)

Application of the Hebbian Learning rule to the passive avoidance paradigm would generate specific predictions regarding rates of learning. Essentially, the pairing of a CS with reward or punishment would be hypothesized to increase the strength of the connection between the unit representing the stimulus and the unit representing the reward or punishment. As such, it would be expected that *the degree to which the reward or punishment units would be active is a function of the degree to which the individual receives reward or punishment.* Following the IES model (Blair, 2004) this would predict that higher rewards and punishments would activate the corresponding units to a greater degree, and

should therefore initiate faster learning, than would lower rewards and punishments. With respect to the purported deficit in passive avoidance learning in individuals with psychopathy, it may be suggested that the disorder is related to weaker activation of the reinforcement units by a given reward or punishment than would be found in a comparison individual (Blair, 2004).

A task manipulation aiming to test this position was incorporated into the passive avoidance task. Thus each CS+ was associated with a different (positive) point value, and each CS- was associated with a different (negative) point value. This task manipulation allowed a sensitive test of two contrasting sets of predictions regarding the proposed stimulus-reinforcement deficit in psychopathy. The first, most parsimonious position, suggests that there exists a general impairment in stimulus-reinforcement learning. This would predict impairment in the formation of both stimulus-reward and stimulus-punishment associations. Indeed, recent data has provided partial support for this view. Adult individuals with psychopathy have presented with impairment in punishment-related emotional learning, while reward-related learning was impaired, but to a lesser degree (Peschardt et al., submitted). The second position, on the basis of previous studies of passive avoidance learning in adult psychopaths, would predict that impairment in emotional learning is confined to aversive stimuli. Indeed, previous studies have suggested that adult individuals with psychopathy approach CS+s at rates comparable to controls (Kosson et al., 1990; Newman and Kosson, 1986; Newman et al., 1990; Thornquist and Zuckerman, 1995). However, the case is unclear as regards children with psychopathic tendencies (Newman et al., 1985; Scerbo et al., 1990).

2.2.2: Summary of Aims

There were two main aims of experiment 1. Firstly, to determine whether the psychopathic tendencies group would show impairment in passive avoidance learning. Specifically, experiment 1 aimed to investigate whether any group differences would be significantly related to psychopathic tendencies after the variation due to level of ADHD has been taken into account. Secondly, to

determine whether the two groups would be differentially affected by varying the level of reward/ punishment associated with individual stimuli.

2.3: Methods

2.3.1: Design

The independent variables were: group (psychopathic tendencies/ comparisons); and level of reinforcement (ranging from –2000 to +2000 points; see section *2.3.4.* for details). Passive avoidance and omission error rates were measured as the dependent variables.

2.3.2: Participants

The participants were all boys aged between 8 and 16 years recruited from three UK government-run schools for children with emotional and behavioural difficulties. They had all received statements under the Education Act of 1993 as being too problematic for mainstream education. All boys taking part in the experiment were Caucasian.

Participants were selected on the basis of the combined APSD scores of two raters (usually two teachers or a teacher and a classroom assistant). In line with previous work (Blair et al., 2001a; Fisher and Blair, 1998), participants with an APSD score of 27 or above were eligible for the psychopathic tendencies group and participants with an APSD score of 15 or below were eligible for the comparison group. Of the boys available for participation, one boy with psychopathic tendencies declined the invitation to participate. Nineteen boys were included in the psychopathic tendencies group and 23 boys were included in the comparison group. It was made clear to all participants that they were free to withdraw from the study at any time.

2.3.3: Measures

British Picture Vocabulary Scale (BPVS; Dunn et al., 1982).

The BPVS was used to measure the participants' verbal intelligence quotient (IQ). The BPVS measures receptive vocabulary for standard English.

Antisocial Process Screening Device (APSD; Frick and Hare, 2001)

Participant's scores for each item were the averages assigned by the two raters. Pearson's correlations of the two ratings for each child were $r2 = 0.77$ ($P < 0.001$) for total APSD score. Inter-rater correlations for the three factors were: $r2 = 0.54$ ($P < 0.001$) for callous/unemotional; $r2 = 0.76$ ($P < 0.001$) for narcissism, and; $r2 = 0.13$ (*NS*) for impulsivity).

ADHD Rating Scale-IV (DuPaul et al., 1998)

Participant's scores for each item were the averages assigned by the two raters. The Pearson's correlation of the two ratings was $r2 = 0.67$ ($P < 0.001$) for total DuPaul score. Inter-rater correlations for the two factors were $r2 = 0.65$ ($P < 0.001$) for hyperactivity-impulsivity and $r2 = 0.66$ ($P < 0.001$) for inattention.

2.3.4: Passive Avoidance Learning Task

The passive avoidance task was a modified version of Newman and Kosson's task (Newman and Kosson, 1986). Stimuli were eight different two-digit numbers which were assigned values of plus-/minus- 1, 700, 1400, or 2000 points. The numbers were controlled for being odd/even and being above or below fifty in order that no attribute could be differentially associated with reward or punishment. The stimuli were also counterbalanced so that those associated with punishment for half of the participants would be associated with reward for the other half. Each stimulus was presented 10 times leading to a total of 80 trials. Trials within each block were presented in a randomised order. Participants had to learn by trial-and-error to press the spacebar key upon presentation of CS+s and to refrain from responding to CS-s. Stimuli remained visible until a response was

made, for up to a maximum of 3 seconds. If a response was made the stimulus disappeared and a message immediately appeared in the centre of the screen indicating how many points had been won or lost (see figure *2.2.* for a diagram depicting a passive avoidance learning trial). If no response was made no feedback was received and the points total remained the same. The stimuli were 2 cm high, presented in white on a black screen. All participants were allocated 10,000 points at the start of the test, and a running total was visible during the feedback display only.

2.3.5: Procedure

Each participant was tested individually in a quiet room allocated by the school. Subsequent to the administration of the BPVS by the experimenter, the participants completed the passive avoidance learning task. The experiment was described without informing the participant of the investigation's specific objectives and expectations. Participants were given the following instructions presented on the computer screen and read aloud by the experimenter – '*In this task, you are going to be presented with a series of numbers. Some of these numbers are good and will gain you points if you press the button when they are showing. Some are bad and will lose you points if you press the button when they are showing. If you do nothing you will neither gain nor lose points. Try to win as many points as you can.*'

figure 2.2: An example of a passive avoidance learning trial

Until a response has been made, or up to 3000 ms

[RESPONSE]

1000 ms

In this trial the participant approached a CS- and was punished with the loss of 2000 points

2.4: Results

As expected one-way ANOVAs showed significant differences between groups in terms of APSD score ($F(1,41) = 404.53$, $P < 0.001$) and DuPaul ADHD score ($F(1,41) = 21.97$, $P < 0.001$). As can be seen in table *2.1.* children with psychopathic tendencies received higher scores on both measures. No significant differences were found between groups in terms of age ($F(1,41) = 1.88$, *NS*) or estimated verbal IQ ($F(1,41) < 1$, *NS*) (see table *2.1.* for full participant details). In order to control for level of ADHD, DuPaul ADHD score was included as a covariate in the analysis.

table 2.1: Mean age, BPVS score, and APSD and ADHD ratings (standard deviations in parentheses)

Group	Age	BPVS	APSD	ADHD
Psychopathic tendencies group (n=19)	**12.31 (2.19)**	**88.58 (13.61)**	**30.01* (2.70)**	**32.53* (9.11)**
Comparison group (n=23)	**13.29 (2.38)**	**88.70 (19.74)**	**11.60 (3.14)**	**17.43 (11.32)**

*** group differences significant at $P < 0.001$**

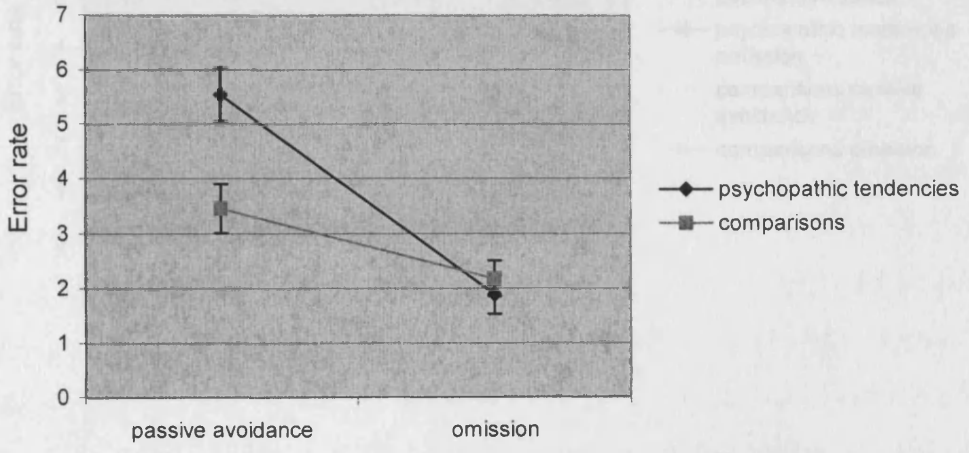
Key to table 2.1: APSD = Antisocial Process Screening Device (maximum score = 40); BPVS = British Picture Vocabulary scale; ADHD = attention-deficit hyperactivity disorder (maximum score = 54); n = number of participants.

Following Newman and Kosson (Newman and Kosson, 1986) each initial presentation of a stimulus was treated as a learning trial, so the first block of trials were excluded from analysis. Data were scored as passive avoidance and omission errors (see figure *2.1.*). A 2 (Group) x 2 (Error) x 4 (Level) mixed model ANCOVA was performed, with DuPaul ADHD score included as the covariate.

The ANCOVA revealed a significant main effect for group ($F(1,39) = 3.91$, $P < 0.05$, 1-tailed). There was also main a effect for error ($F(1,39) = 4.95$, $P < 0.05$) and a significant interaction between group and error ($F(1,39) = 4.71$, $P <$

0.05). Neither the main effect of the covariate ($F(1,39) < 1$, *NS*) nor any interactions involving the covariate were significant. As can be seen in figure *2.3.*, the main effects involving group, and the interaction between group and error were driven by the group difference in passive avoidance errors: whilst the children with psychopathic tendencies made more passive avoidance errors, the level of omission errors was similar between groups. Figure *2.3.* also clearly shows that the main effect of error was due to an increase, by both groups, in number of passive avoidance rather than omission errors.

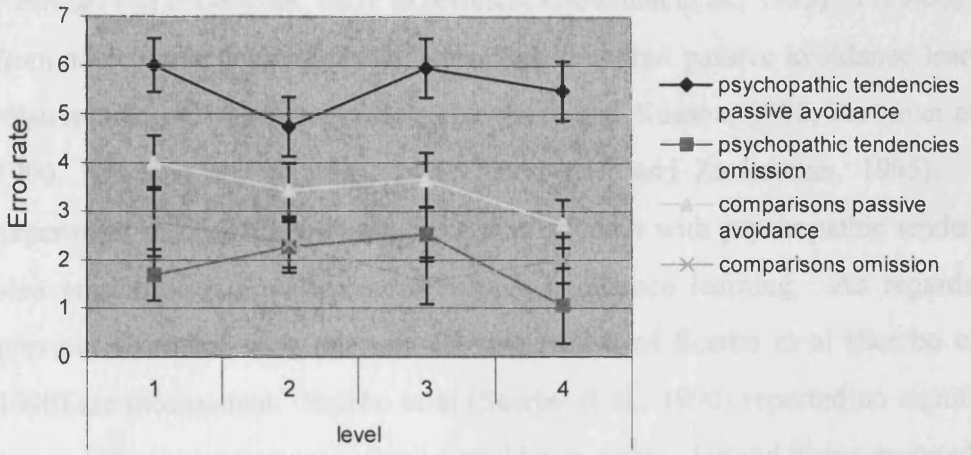
figure 2.3: Passive avoidance and omission errors by group



Due to the non-significant effects produced by the covariate, the analysis was repeated excluding the covariate. As previously, the ANOVA revealed main effects for group ($F(1,40) = 7.05$, $P < 0.01$) and error ($F(1,40) = 27.23$, $P < 0.001$), and also an interaction between group and error ($F(1,40) = 6.29$, $P < 0.01$) (see figure *2.3.*). Additionally, the extra power led to a significant 3-way interaction, between group, error and level of points ($F(3,120) = 2.33$, $P < 0.05$, 1-tailed) (see figure *2.4.*). Follow-up tests were performed for both groups and errors separately. The main effects were neither significant for level of punishment ($F(3,54) = 2.33$, *NS*) nor level of reward ($F(3,54) = 2.04$, *NS*) for the children with psychopathic tendencies. For the comparison children, however, there was a significant main effect of level of punishment on number of passive

avoidance errors committed ($F(3,66) = 2.43$, $P < 0.05$, 1-tailed). Further, a significant linear effect ($F(1,22) = 9.27$, $P < 0.01$) demonstrated that the comparison children made fewer passive avoidance errors as the point value associated with the CS- increased (see figure *2.4.*). As regards the omission error data, there was no significant effect of level of reward for the comparison group ($F(3,66) < 1$, *NS*).

figure 2.4: Passive avoidance and omission errors by value



2.5: Discussion

The main aim of experiment 1 was to investigate passive avoidance learning ability in boys with psychopathic tendencies. Specifically it aimed to investigate whether any group differences were significantly related to psychopathic tendencies after the variation due to level of ADHD had been removed. A further aim was to assess any effects of varying the level of reward and punishment associated with individual stimuli.

In line with predictions, the psychopathic tendencies group made more passive avoidance errors than the comparison group, moreover, the effects were still present after co-varying level of ADHD. As regards level of punishment, the results looked similar to those recently obtained with adult psychopathic individuals (Blair et al., 2004). The children with psychopathic tendencies were

less sensitive to the degree of punishment than the comparison boys. Further, the number of passive avoidance errors committed by comparison children decreased in a linear fashion as the negative point value associated with the stimuli increased. This effect was not observed in the psychopathic tendencies group. As expected there were no group differences according to omission errors. However, contrary to expectations, the performance of neither group was significantly modulated by level of reward.

The results of experiment 1 are in line with the adult literature and Newman and colleagues' early experiment (Newman et al., 1985). Previous data from adults have consistently demonstrated impaired passive avoidance learning relative to comparison individuals (Newman and Kosson, 1986; Newman et al., 1990; Newman and Schmitt, 1998; Thornquist and Zuckerman, 1995). This experiment provides further evidence that children with psychopathic tendencies also present with impairment in passive avoidance learning. As regards the previous literature, it is unclear why the results of Scerbo et al (Scerbo et al., 1990) are inconsistent. Scerbo et al (Scerbo et al., 1990) reported no significant group differences in rates of passive avoidance errors. Instead it was reported that the adolescents with psychopathic tendencies made significantly fewer omission errors than the comparison adolescents. One possibility that may explain this inconsistency concerns participant selection. Scerbo et al (Scerbo et al., 1990) used a self-report measure of psychopathy which may have led to a heterogenous population of children. A recent study reported that children with bipolar disorder made fewer omission errors than comparison children (Gorrindo et al., in press). Indeed the two disorders share some similar identification criteria, such as grandiosity, inflated self-esteem and distractibility (see table *1.2.*) (APA, 1994). Thus, it could be that at least some of the 'psychopathic delinquents' in the Scerbo et al study were actually suffering from bipolar disorder.

As regards ADHD, there were no significant effects involving this covariate. Two previous studies, however, have provided contradictory results, suggesting that passive avoidance learning is impaired in individuals with ADHD (Hartung et al., 2002; Iaboni et al., 1995; Milich et al., 1994). However, it should

be noted that Iaboni et al (Iaboni et al., 1995) did not examine level of psychopathic tendencies, and further, children were included in the study regardless of conduct disorder (CD) diagnosis (although the presence of CD was not taken into account in analyses). In the current experiment, there would have been significant differences in the groups divided by their level of ADHD symptomatology (if level of APSD score had not been co-varied). The results of experiment 1 are in line with the results of Hartung et al (Hartung et al., 2002). While they found that ADHD symptoms in boys and girls were predictive of passive avoidance errors, a hierarchical regression analysis indicated that passive avoidance learning ability was related most strongly to CD (a diagnosis that most children with psychopathic tendencies would meet).

Finally, it must be noted that, contrary to expectations, there were no effects, in either group, regarding the level of reward. There were, however, very few errors of omission made by either group, thus there may have been ceiling effects. Level of reward could be tested in the future, possibly by increasing the number of CS+s in the task. Such a modification may increase task load, therefore reducing ceiling effects. Alternatively the task used in experiment 1 may not have been sensitive enough to test for level of reward. This may be investigated by altering the incentive values associated with CS+s. In keeping with these suggestions, there were effects of reward value observed in a modified version of this task that involved a greater number of stimuli using adult psychopaths (Blair et al., 2004).

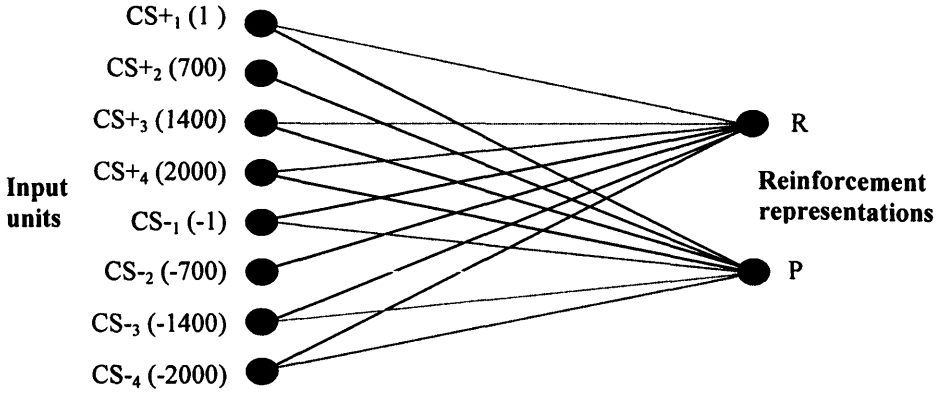
2.6: Summary and Conclusions

In conclusion, the present experiment has replicated previous findings with psychopathic adults indicating that this neuro-cognitive disorder is also associated with poor passive avoidance learning in childhood. Further, passive avoidance error rates were modulated in comparison children by modifying the incentive value of the CS-s. This effect was not observed in children with psychopathic tendencies.

2.7: Experiment 2

Experiment 1 demonstrated that there exists passive avoidance learning impairment in children with psychopathic tendencies. Experiment 2 aimed to further characterise the impairment by comparing the data from experiment 1 with a simple connectionist simulation of passive avoidance learning.

figure 2.5: A connectionist model depicting successful passive avoidance learning



The model portrays the relative weights between the input units (CS+1 – CS+4 and CS-1 – CS-4) and the reinforcement representation units (R and P). Red connections indicate the highest weights, followed by orange, yellow and pink. Black connections designate low weights (*i.e.* minimal association) between a stimulus and the unit representing reward/punishment.
Key, R=reward, P=punishment, values in parentheses following input units indicate hypothetical point values associated with the units following the task from experiment 1.

As discussed above, it may be hypothesized that passive avoidance learning occurs during a process akin to Hebbian Learning (see figure *2.5.*). The basic suggestion implicit in this model is that performance on the passive avoidance learning task requires the participant to form appropriate stimulus-reinforcement associations such that the individual approaches CS+s (stimuli associated with reward) and avoids CS-s (stimuli associated with punishment) (see section *2.1.*). The pairing of a stimulus with reward or punishment would

increase the strength of the connection between the unit representing the stimulus and the unit representing the reward or punishment through Hebbian Learning (Hebb, 1949). In other words activation of the reward unit initiates approach behaviour, whilst the converse occurs upon activation of the punishment unit.

To illustrate, (see figure *2.5.*) a response to $CS+_1$ will result in reward. In terms of the model, the unit representing $CS+_1$ and the reward (R) unit will be simultaneously active and the strength of the connection between these units will increase as a function of the level of activation of the two units (Hebb, 1949) (see section *2.2.1.* for the Hebbian Learning Rule). Moreover, as indicated above, the degree to which the reward or punishment units are active is a function of the degree to which the individual receives reward or punishment; higher rewards and punishments activate the corresponding units to a greater degree than do lower rewards and punishments. In short, the connection between the unit representing $CS+_4$ and the reward unit ought to be stronger than that between the unit representing $CS+_1$ and the reward unit.

Following the positions outlined in experiment 1 regarding the nature of the proposed stimulus-reinforcement impairment in psychopathy, the computational model aimed to simulate passive avoidance learning under: (1) normal conditions (*i.e.* an attempt to simulate the performance of comparison individuals), (2) conditions of general emotional learning impairment (*i.e.* impaired stimulus-reward *and* stimulus-punishment associations) and finally, (3) conditions of impairment exclusive to the learning of stimulus-punishment associations.

2.7.1: Summary of Aims

Experiment 2 aimed to develop a connectionist model of passive avoidance learning in order to further characterize the impairment shown by children with psychopathic tendencies in experiment 1.

2.8: Methods

2.8.1: Connectionist Modelling

The model was based on the task from experiment 1, thus there were eight input units (four for each of the CS+s and four for each of the CS-s) and two reinforcement representations (reward and punishment) (see figure *2.5.*). The input units were considered to be activated by the presentation of a stimulus. Activation of the reinforcement representations was determined by the following simple formula (see formula *2.1.*):

Formula 2.1: $$a_r = a_i \,.\, w(a_i.a_r)$$

where a_r is the activation of the reinforcement (reward or punishment) unit, a_i is the activation of the input unit and $w(a_i.a_r)$ is the weight of the connection between the input unit and that reward or punishment unit.

The decision to respond or not was a function of the degree to which the reward or punishment unit was activated (by the associated weight) (see formula *2.2.*):

Formula 2.2: $$p(\text{respond}) = a_r/(a_r + a_p)$$

where a_r is the activation of the reward unit and a_p is the activation of the punishment unit.

At the beginning of learning, the weights between the inputs units and the reward unit were set at 0.7 and the weights between the input units and the punishment unit were set at 0.3; *i.e.*, the model was initially biased to respond to stimuli[2].

The model experienced each of the eight stimuli over the course of 10 blocks of trials (following the task procedure described in experiment 1). If the model 'decided' to respond to a stimulus, it received reward (if the stimulus was a

[2] This bias was introduced in order to simulate initial rates of approach and avoidance behaviour in the first block of trials by both groups.

CS+) or punishment (if the stimulus was a CS-) *i.e.*, the reward or punishment units would be activated. Following the receipt of the feedback, learning would occur; *i.e.*, the weights between the input unit and the reward or punishment units were modified as a function of Hebbian Learning (Hebb, 1949) (see formulae *2.1.* and *2.2.*).

The degree to which the reward or punishment units were activated was a function of the level of points gained or lost using the prospect theory value function (Kahneman and Tversky, 1979). Essentially this value function states that gains and losses are evaluated relative to reference points rather than absolute value. Further, it states that losses are 'disliked' about twice as much as absolute equivalent gains are 'liked'. The resulting value function is thus steeper for losses than gains.

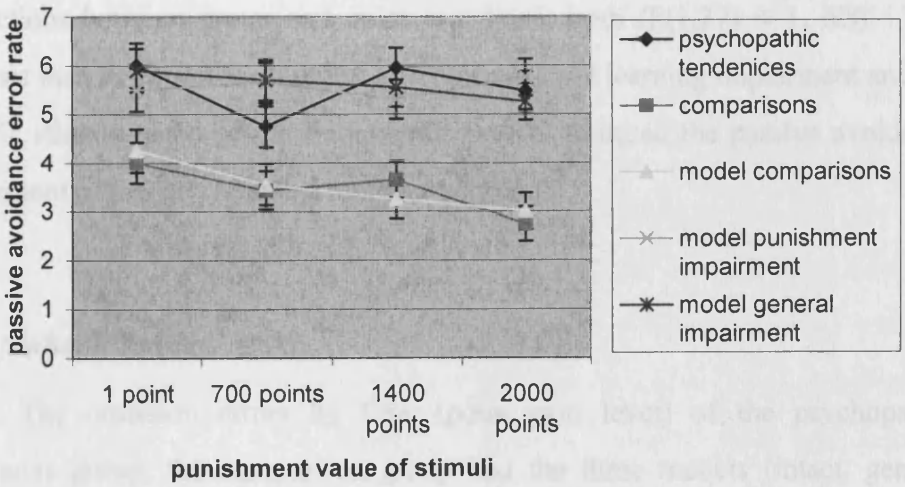
The two alternative predictions regarding the performance of the psychopathic tendencies group (*i.e.* that there exists a general emotional learning impairment or that there exists a specific stimulus-punishment impairment) were modelled by assuming that either (i) there was general hypo-responsiveness of the reward and punishment representations in individuals with psychopathy (*i.e.*, both reward and punishment units were 90% less active), or (ii) by assuming that there was only hypo-responsiveness of the punishment representations (*i.e.*, only the punishment unit was 90% less active). Data generated by these two contrasting computational descriptions of psychopathy, together with the predictions of the model for comparison individuals (*i.e.* both reward and punishment units were functioning at 100%), were then compared against the actual data presented in experiment 1.

2.9: Results

2.9.1: Passive Avoidance Errors

Figure *2.6.* depicts the passive avoidance errors by CS- (point loss level) of the psychopathic tendencies group, the comparison group and the three models (intact, general emotional learning impairment and specific stimulus-punishment impairment). Initially, the model of the comparison passive avoidance learning was compared with the data from the comparison children using a 2 (Group; comparison group vs. intact model) x 4 (Level of point loss) ANOVA. There was a significant main effect of point loss level ($F(3, 123) = 6.17$; $P < 0.001$). However, as predicted neither the main effect of group ($F(1, 41) < 1$; *NS*) nor the interaction between group and point loss level ($F(3, 123) < 1$; *NS*) were significant. In short, the intact model successfully captured the passive avoidance learning performance of the comparison children.

figure 2.6: Passive avoidance errors made by the children and the connectionist models



In order to add validity to the comparison model of passive avoidance errors it was necessary to directly compare it with the general emotional learning impairment and specific stimulus-punishment impairment models. For this purpose, two further ANOVAs were conducted (2 Group x 4 Level ANOVA).

These revealed large group effects ($F(1, 41) = 15.97$ & 16.83 respectively; $P < 0.001$). In short, as expected, the performance of neither the general emotional learning impairment nor specific stimulus-punishment impairment models described the passive avoidance learning of the comparison children accurately.

As regards the psychopathic tendencies group, the first analysis aimed to confirm that the model of the comparison individuals did *not* capture their passive avoidance performance. A 2 (Group; psychopathic tendencies group vs. intact model) x 4 (Level of points loss) ANOVA was conducted on the data. This revealed a highly significant group difference ($F(1, 37) = 19.28$; $P < 0.001$). In short, in line with predictions, the intact model did not capture the performance of the psychopathic tendencies children.

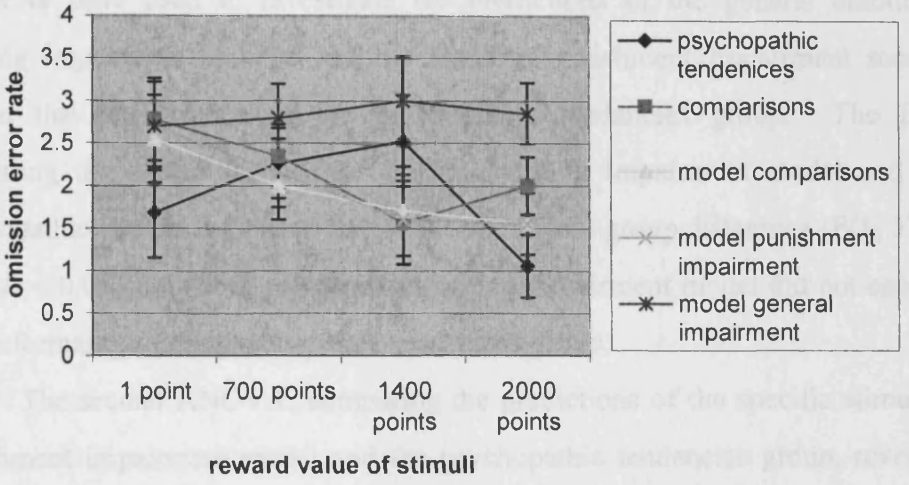
Following this, it was necessary to examine whether the general emotional learning impairment and specific stimulus-punishment impairment models successfully captured the passive avoidance performance of the psychopathic tendencies group. Two 2 (Group) x 4 (Level of points loss) ANOVAs were performed. Neither of these revealed significant group differences or significant interactions between group and point loss level; both ($F(1,37) < 1$, *NS*). This suggests that, as expected, both the general emotional learning impairment and the specific stimulus-punishment impairment models captured the passive avoidance impairment of the psychopathic tendencies group.

2.9.2: Omission Errors

The omission errors by CS+ (point gain level) of the psychopathic tendencies group, the comparison group and the three models (intact, general emotional learning impairment, specific stimulus-punishment impairment) are depicted in figure *2.7.* Again an initial analysis was performed in order to confirm that the model of the comparison individuals captured the pattern of omission errors of the comparison children. A 2 (Group; comparison group vs. intact model) x 4 (Level of points gain) ANOVA was performed. Neither the main effect of group ($F(1, 41) < 1$; *NS*) nor the interaction between group and point loss

(F(3, 123) < 1; *NS*) were significant. However, there was a significant linear effect of point level (F(1, 41) = 3.42; $P < 0.05$, 1-tailed). This suggests that the intact model did capture the performance of the comparison individuals successfully. As can be seen in figure *2.7.* the performance of both groups was modulated by level of reward; with increasing reward leading to decreasing errors.

figure 2.7: Omission errors made by the children and the three connectionist models



The general emotional learning impairment model assumes that the formation of stimulus-reward associations is impaired. This model should not therefore capture the performance of the comparison group. To test this prediction, a 2 (Group) x 4 (Level of point gain) ANOVA comparing the data form the comparison group to the predictions of the general emotional learning impairment model was performed on the data. This revealed only a group effect (F(1, 41) = 3.09; $P < 0.05$, 1-tailed). In short, general emotional learning impairment model did not describe the omission data generated by the comparison group.

The specific stimulus-punishment impairment model assumes no impairment in the formation of stimulus-reward associations and thus should describe the omission data for the comparison individuals. Indeed, a 2 (Group) x 4 (level of point gain) ANOVA comparing the data form the comparison group to

the predictions of this model revealed no group effect or group by level of point gain interaction ($F(1, 41) < 1$ for both; *NS*). There was, however, again a significant linear effect of point level ($F(1, 41) = 4.44$; $P < 0.05$).

The general emotional learning impairment and the specific stimulus-punishment impairment models differed in their predictions for the omission data of the psychopathic tendencies group, with the former predicting impairment and the latter predicting intact performance. Two 2 (Group) x 4 (Level of points gain) ANOVAs were used to investigate the predictions of the general emotional learning impairment and the specific stimulus-punishment impairment models against the performance of the psychopathic tendencies group. The first, comparing the predictions of the general learning impairment model and the psychopathic tendencies group, revealed a significant group difference ($F(1, 37) = 5.94$; $P < 0.05$). In short, the general learning impairment model did not capture the performance of the psychopathic tendencies group.

The second ANOVA, comparing the predictions of the specific stimulus-punishment impairment model and the psychopathic tendencies group, revealed no significant group effect ($F(1,37) < 1$, *NS*). There were, however, trends towards a significant main effect of point level ($F(3,111) = 2.27$, $P < 0.085$) and a significant interaction between group and points level ($F(3,111) = 2.42$, $P < 0.07$). As is apparent in figure *2.7.*, this indicates that the specific stimulus-impairment model did not *fully* fit the omission error data of the children with psychopathic tendencies.

2.10: Discussion

Experiment 2 described the development of a computational account of passive avoidance learning to simulate performance of the participants in experiment 1. The intact model was designed to replicate the performance of the comparison group. It was then impaired in two contrasting ways and performance of these models was compared with the performance of the children with psychopathic tendencies presented in experiment 1. The model was impaired in

order to simulate: (1) a general emotional learning impairment (*i.e.* impaired ability to form both stimulus-reward and stimulus-punishment associations); and (2) a specific punishment-learning impairment (*i.e.* impaired ability to form stimulus-punishment associations, but intact ability to perform stimulus-reward associations).

The computational simulation of the passive avoidance task assumed that participants decided whether to respond to a CS as a function of the formation of stimulus-reinforcement associations. As can be seen in figures *2.6.* and *2.7.*, the model very successfully described the performance of the comparison group. In particular, it modelled, relatively successfully, the impact of level of reward and punishment on the pattern of errors made.

As regards the two possible models of psychopathy, the data obtained in experiment 1 matched the predictions of the selective impairment for stimulus-punishment associations in terms of both overall error rate and pattern of errors produced (according to level of punishment). In terms of omission errors, the specific stimulus-punishment impairment model simulated the data of the children with psychopathic tendencies more successfully than the general emotional learning impairment model. A glance at figure *2.7.*, however, indicates that while there were no overall differences in omission error rates between the children with psychopathic tendencies and the selectively impaired model, the *pattern* of data was not consistent with the idea that stimulus-reward learning is entirely unimpaired. This result is consistent with a recent report, indicating that, while to a lesser degree than stimulus-punishment learning, stimulus-reward learning is also impaired in individuals with psychopathy (Peschardt et al., submitted).

2.11: Summary and Conclusions

Experiment 2 tested a Hebbian Learning simulation of passive avoidance learning against the data collected in experiment 1. The model captured the performance of the comparison children very successfully in terms of both passive avoidance and omission error data. As expected, the passive avoidance errors made by the children with psychopathic tendencies were modelled successfully by

both the selectively impaired stimulus-punishment model and the general learning-impairment model. The pattern of omission errors produced by the children with psychopathic tendencies was not captured entirely successfully by either the assumption that stimulus-reward learning is either entirely intact, or impaired to the same level as stimulus-punishment associations.

2.12: General Discussion

Experiment 1 demonstrated that, in line with the adult literature, children with psychopathic tendencies showed impairment in passive avoidance learning relative to comparison children. In addition, this impairment was observed even after controlling for level of ADHD. Experiment 2 compared the data from experiment 1 with a connectionist simulation of passive avoidance learning. The impaired models successfully captured the pattern of passive avoidance learning made by children with psychopathic tendencies.

2.12.1: Implications of these Results for the Characterization of Psychopathy

The results of experiments 1 and 2 indicate that children with psychopathic tendencies are impaired in passive avoidance learning. This is in line with previous data collected with adult psychopaths (Newman and Kosson, 1986; Newman et al., 1990; Newman and Schmitt, 1998; Thornquist and Zuckerman, 1995) and also one previous report of passive avoidance learning impairment in adolescents with psychopathic tendencies (Newman et al., 1985). Following Blair (2004), these data, assuming an amygdala-based pathology, are also in line with previous data indicating that both adult psychopaths and children with psychopathic tendencies perform similarly poorly on other tasks hypothesized to be reliant upon the integrity of the amygdala, such as recognition of and emotional responsivity to fearful and sad affect (Aniskiewicz, 1979; Blair, 1999; Blair et al., 1997; House and Milligan, 1976; Sutker, 1970; Blair et al., 2001c; Blair and Coles, 2000; Stevens et al., 2001).

2.12.2: Impact of ADHD

As noted above, the group differences in passive avoidance learning were significant even after controlling for ADHD. This is in keeping with the idea, again assuming that passive avoidance is an amygdala-reliant form of emotional learning, that ADHD is not associated with amygdala dysfunction. Indeed, ADHD has been previously associated with 'executive dysfunction', and in particular, with impairment in 'response inhibition' (Barkley, 1999). It could be suggested that passive avoidance requires behavioural inhibition as the participant must inhibit responses to stimuli associated with punishment. However that ADHD was not a significant covariate in the analyses in experiment 1, suggests that the form of behavioural inhibition which, when impaired, may lead to ADHD is dissociable from the process that is involved in passive avoidance learning. Further, passive avoidance learning performance was modelled relatively successfully in experiment 2 in the absence of ADHD, that is, ADHD was not included in the connectionist model.

2.12.3: Implications of these Results for the Theories of Psychopathy

The current results are relevant for the fear dysfunction hypotheses, the Response-Set Modulation (RM) hypothesis and the Integrated Emotion Systems (IES) hypothesis. As regards passive avoidance learning, the fear positions would predict impairment owing to an insensitivity to punishment information. However, it must be noted that the *pattern* of performance on the *rewarded* trials was also abnormal in children with psychopathic tendencies. In contrast, the pattern of performance of the comparison children was predicted relatively successfully by the connectionist model, both in terms of reward- and punishment-related processing; in both cases there was a linear relationship between reinforcement and error rate. In contrast, the pattern of performance of the children with psychopathic tendencies was only partially successfully predicted by the specific stimulus-punishment impairment model; the specific stimulus-punishment account did not predict the abnormal pattern of performance as regards omission errors. In short, it appears that the ability to modulate

performance as a function of punishment, but also to a lesser degree, reward, is impaired in children with psychopathic tendencies.

The RM model predicts that individuals with psychopathy will be more likely than comparison individuals to respond even when a salient punishment stimulus (*i.e.* a CS-) is present on the screen during tasks of passive avoidance learning. As regards the RM hypothesis, similar criticisms as were suggested above apply. Namely, that there was also evidence of an abnormal pattern of performance in the reward-related trials, thus an explanation focusing on punishment is essentially inadequate. Further, according to the model, *"the impulsivity, poor passive avoidance, and emotion-processing deficits of individuals with psychopathy may all be understood as a failure to process the meaning of information that is peripheral or incidental to their deliberate focus of attention"* (Lorenz and Newman, 2002). However, as noted in section *1.3.5.* in standard tasks of passive avoidance learning (as in the task used in experiment 1) the punishment information appears on screen *in the absence of any competing information.* Thus it is unclear how this attentional explanation can account for the deficit.

Finally the IES model (Blair, 2004) predicted impairment in forming stimulus-punishment associations in this population. Further, it also predicts that there may too be impairment in the formation of stimulus-reward associations (as these are also hypothesized to be amygdala-reliant). Thus this position is successfully able to account for the results from chapter 2. The results here (as suggested above in section *2.12.1.*) also therefore strengthen claims made by the IES position that there exists amygdala impairment in children with psychopathic tendencies.

2.13: Conclusions

Chapter 2 demonstrated that children with psychopathic tendencies are impaired in passive avoidance learning. The following chapter will investigate the neural substrates involved in successful passive avoidance learning. This may

serve to identify neural regions that might be dysfunctional in individuals with psychopathy.

Chapter 3 – The Neural substrates of Passive Avoidance Learning in Healthy Adults

3.1: Experiment 3

As demonstrated in chapter 2, and previous experimental work (Blair et al., 2004; Newman et al., 1990; Newman and Schmitt, 1998; Newman et al., 1985), children with psychopathic tendencies and adult psychopaths present with impairment in passive avoidance learning. Experiment 3 aims to identify the neural substrates underlying passive avoidance learning in healthy adults. This may serve to identify neural structures as targets for future research efforts with this population.

As noted in section *2.1.* passive avoidance learning is thought to require the formation of stimulus-reward and stimulus-punishment associations (Baxter and Murray, 2002). Electrophysiological, lesion, and pharmacological intervention studies with animals have indicated that passive avoidance learning is reliant upon a network of neural structures including amygdala, orbitofronal cortex (OFC), insula, striatum and hippocampus. Lesions and pharmacological interventions involving the amygdala (Ambrogi Lorenzini et al., 1991; Bermudez-Rattoni et al., 1997; Bermudez-Rattoni and McGaugh, 1991; Cahill and McGaugh, 1990; Treit and Menard, 1997), insula (Bermudez-Rattoni et al., 1997; Bermudez-Rattoni et al., 1991; Bermudez-Rattoni and McGaugh, 1991; Gutierrez et al., 1999), striatum (Ambrogi Lorenzini et al., 1995; Sandberg et al., 1984) and hippocampus (Ambrogi Lorenzini et al., 1997; Gray, 1987; McGaugh, 2002) have all impaired passive avoidance learning. Further, animal work has suggested that amygdala is crucial for learning associations between sensory stimuli and cues for incentive stimuli attributes (Blundell et al., 2003). In addition, single cell recording work with rodents has demonstrated that neurons in the OFC, striatum and basolateral amygdala come to show selective responding during odour-cue passive avoidance learning (Schoenbaum et al., 1999). Specifically, selective neurons in these regions show increased activity to odours predictive of reward

while other neurons show increased activity to odours predictive of punishment (Schoenbaum et al., 1998; Schoenbaum et al., 1999; Setlow et al., 2003). Interestingly, while neurons in OFC do show selective responding during passive avoidance tasks (Gallagher et al., 1999; Schoenbaum et al., 1998; Tremblay and Schultz, 2000), these neurons do not appear to be *necessary* for successful behavioural performance in these tasks. Indeed, lesions of OFC have left ability to perform passive avoidance learning intact in rodents (Schoenbaum et al., 2002).

Surprisingly, there have been no functional imaging studies of passive avoidance learning in humans. Indeed, there have been relatively few investigations of the neural systems involved in instrumental learning in humans. In one of the few existing studies Elliott and colleagues (Elliott et al., 2004) reported that amygdala activation was enhanced when participants made instrumental responses for reward. In contrast error-related instrumental learning has been associated with activity in several striatal regions including the caudate, nucleus accumbens and putamen (Elliott et al., 2004; O'Doherty et al., 2004; Pagnoni et al., 2002). In addition, and consistent with suggestions that OFC plays an important role in incentive valuation (Tremblay and Schultz, 1999), Elliott and colleagues (Elliott et al., 2004) reported OFC and rostral ACC activation to rewarding stimuli whether these stimuli required an instrumental response to obtain reward or not. Further, other work (*e.g.* Cox et al., 2005; Elliott et al., 2000a) has also implicated medial OFC/ rostral cingulate in the representation of reward.

3.1.1: Summary of Aims

This experiment aimed to determine the BOLD responses associated with passive avoidance learning in healthy human participants. Following the animal passive avoidance literature and previous FMRI results from human instrumental learning paradigms, it was predicted that successful passive avoidance learning would be associated with an integrated neural response including medial OFC/ rostral ACC, the insula, striatum, hippocampus and amygdala.

3.2: Methods

3.2.1: Participants

Twenty right-handed adults participated in the study. One of the participants performed the task too well (precluding an adequate number of post-criterion passive avoidance error trials) so the data from nineteen participants (9 women and 10 men, mean age = 30.74; SD = 6.07; range = 22-40 years) were analysed. All participants were in good health with no past history of psychiatric or neurological disease and gave informed written consent.

3.2.2: MRI Data Acquisition

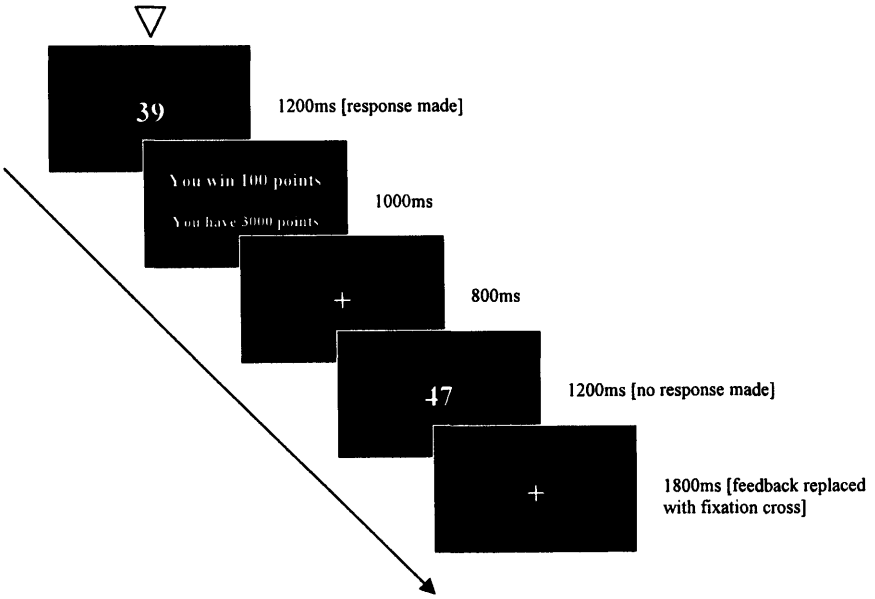
Participants were scanned during task performance using a 1.5 Tesla GE Signa scanner. A total of 207 functional images were taken per run with a gradient echo echo-planar imaging (EPI) sequence (repetition time = 2500 ms, echo time = 40 ms, 64 x 64 matrix, flip angle 90°, FOV 24 cm). Whole brain coverage was obtained with 29 axial slices (thickness, 4-mm; in-plane resolution, 3.75 x 3.75 mm). A high-resolution anatomical scan (three-dimensional Spoiled GRASS; repetition time = 8.1 ms, echo time = 3.2 ms; field of view = 24 cm; flip angle = 20°; 124 axial slices; thickness = 1.0 mm; 256 x 256 matrix) in register with the EPI dataset was obtained covering the whole brain.

3.2.3: Passive Avoidance Task and Experimental Procedure

Experiment 3 involved an event-related design. Participants completed a total of three runs, each of 8.5 minutes, which were presented in a randomized order. Within each run the participants were presented with a separate version of the passive avoidance task (which differed only in the stimuli used). The participants were placed in a light head restraint within the scanner to limit head movement. Before entering the scanner, participants performed a 10 trial training run with one pair of stimuli to familiarize themselves with the paradigm. The tasks were programmed in E-Studio.

Test stimuli (12 different two-digit numbers) were presented serially, in a randomized order. Six of these stimuli were 'good'; if participants responded to these they received positive (*i.e.* rewarding; *"you win 100 points"*) feedback. The remaining six were 'bad'; if participants responded to these they received negative (*i.e.* punishing; *"you lose 100 points"*) feedback. If the participant chose not to respond to a test stimulus, they neither won nor lost points, and the feedback display was replaced with a fixation point. Trials lasted 3000 ms and involved the presentation of: a fixation cross for 200 ms, the test stimulus for 1200 ms, feedback for 1000 ms and finally a fixation cross for 600 ms (see figure *3.1.*).

figure 3.1: A timeline depicting the passive avoidance learning task



(timeline running from top to bottom)
The white arrow indicates that the participant approached stimulus '39', which led to reward. In the next trial the participant avoided stimulus '47' and the feedback display was replaced with a fixation cross.

Runs comprised 10 consecutive blocks of 16 trials (in addition to the 12 test stimuli, 4 fixation trials were presented per block to serve as a baseline). Each test stimulus was repeated 10 times (once in each of the 10 blocks, the order of the test stimuli was randomised across blocks). At the beginning and end of each run

a fixation cross was displayed for 1500 ms. The stimuli were presented in a white font on a black computer display projected onto a mirror in the MRI scanner. Participants were able to respond, by right thumb button press, only while the test stimulus was present on the screen. Participants began the task with 0 points and a running score of their total points was displayed at the bottom of the screen during the feedback display.

As in the task used in experiment 1, the design yielded two types of correct response (hits and correct rejections) and two types of incorrect response (passive avoidance and omission errors) (see figure *2.1.*). Pilot work indicated that healthy participants showed rapid learning during the first few presentations of individual stimuli, after which their behavioural performance reached a plateau. At this point it may be considered that the participant has learned the task successfully, *i.e.* they have reached criterion performance. On the basis of the behavioural data (see figure *3.2.*), trials in blocks 2-4 were considered pre-criterion, and trials in blocks 5-10 were considered post-criterion. Data from block 1 were modelled separately (as an event of no interest) as this was the participant's first experience with each of the stimuli. Thus, for the purposes of analysis, events were divided into eight types according to; (i) Learning Stage (whether the action occurred pre- or post- learning criterion); (ii) Response Accuracy (whether the action was correct or incorrect); and, (iii) Stimulus Valence (whether the participant's action was to a 'good' or 'bad' stimulus).

These eight event types are: (1) pre-criterion hit; (2) post-criterion hit; (3) pre-criterion correct rejection; (4) post-criterion correct rejection; (5) pre-criterion miss; (6) post-criterion miss; (7) pre-criterion passive avoidance error; and, (8) post-criterion passive avoidance error.

3.2.4: Behavioural Data Analysis

The behavioural data were analysed using SPSS. Rates of incorrect responses were analysed using a 2 (Incorrect Response; approach/avoidance) x 10 (Block) repeated measures ANOVA.

3.2.5: FMRI Analysis

Data were analysed within the framework of the general linear model using Analysis of Functional Neuroimages (AFNI; Cox, 1996). Both individual and group-level analyses were conducted. The first six volumes in each scan series, collected before equilibrium magnetization was reached, were discarded. Motion correction was performed by registering all volumes in the EPI dataset to a volume collected shortly before acquisition of the high-resolution anatomical dataset.

The EPI datasets for each participant were spatially smoothed (using an isotropic 6mm Gaussian kernel) to reduce the influence of anatomical variability among the individual maps in generating group maps. Next, the time series data were normalized (by dividing the signal intensity of a voxel at each time point by the mean signal intensity of that voxel for each run and multiplying the result by 100). Resultant regression coefficients represented percent signal change from the mean. Following this, regressors depicting each of the eight response types (and one regressor of no interest modelling the data of block 1) were created by convolving the train of stimulus events with a gamma-variate haemodynamic response function to account for the slow haemodynamic response (Cohen, 1997). Linear regression modelling was performed using the nine regressors described above plus regressors to model a first order baseline drift function. This produced for each voxel and each regressor, a beta coefficient and its associated t-statistic.

Single subject beta coefficients were transforming into the standard coordinate space of Talairach and Tournoux (Talairach and Tournoux, 1988). Voxel-wise group analyses involved performing two-sample random effects t-tests contrasting the beta coefficients for the following:

(i) post-criterion correct responses vs. all incorrect responses ([post-criterion hits + post-criterion correct rejections]/2) – ([pre-criterion misses + pre-criterion passive avoidance errors + post-criterion misses + post-criterion passive avoidance errors]/4). This contrast was performed specifically to observe activation related to the learned correct responses, that is, to examine the effects of

accuracy irrespective of the stimulus valence (and thus also irrespective of the specific action performed);

(ii) post-criterion correct responses vs. pre-criterion correct responses ([post-criterion hits + post-criterion correct rejections)/2] – ([pre-criterion hits + pre-criterion correct rejections]/2). This contrast was performed to examine the effects of *stage* of learning, that is, to assess changes in activation as learning improved. Again this contrast was interested in effects irrespective of the stimulus valence;

(iii) post-criterion hits vs. post-criterion correct rejections. This contrast was performed in order to examine any different effects produced by *stimulus valence* (and also then was confounded with action).

The contrasts produced group maps of areas of differential activation ($P < 0.001$). To correct for multiple comparisons a spatial clustering operation was performed using AlphaSim (Ward, 2000) with 1,000 Monte Carlo simulations taking into account the entire EPI matrix ($P < 0.05$). Finally, a conjunction analysis of post-criterion hits vs. pre-criterion hits, plus post-criterion correct rejections vs. pre-criterion correct rejections, was also performed (due to reduced power, for this contrast $P < 0.01$). This was essentially the opposite of contrast (iii) above, that is, it examined common effects of learning the *correct action* to a stimulus (regardless of the valence of the stimulus).

A subset of clusters showing significant differential activation within each contrast were selected according to a priori predictions about the regions involved in passive avoidance learning. These clusters were used to define functional regions of interest (ROIs). In addition, a priori predictions about the involvement of the amygdala in passive avoidance learning justified use of an ROI approach to investigate activation within this region. Thus, standardized bilateral ROIs of the amygdala (identified using a pre-defined AFNI template) were applied to the data. Following this, areas of significant differential activation within the template were sampled, resulting in an ROI encompassing only significant activation within the amygdala ($P < 0.05$, uncorrected for multiple comparisons).

Average percent signal change was measured within each ROI, and data were analysed using repeated measures ANOVAs [2 (Stimulus Valence; 'good'/'bad') x 2 (Response Accuracy; correct/incorrect) x 2 (Stage; pre-/post-criterion)].

3.3: Results

3.3.1: Behavioural Data

Figure *3.2.* depicts the number of passive avoidance errors and avoidances across blocks averaged across the three tasks completed (see table *3.1.* for the average number of events). An ANOVA revealed significant main effects for Incorrect Response ($F(1, 18) = 10.76$, $P < 0.001$) and Block ($F(9, 162) = 62.93$, $P < 0.001$), and also interaction between Incorrect Response and Block ($F(9, 162) = 74.22$, $P < 0.001$).

table. 3.1: Event frequencies according to Response Accuracy, Stimulus Valence and Learning Stage

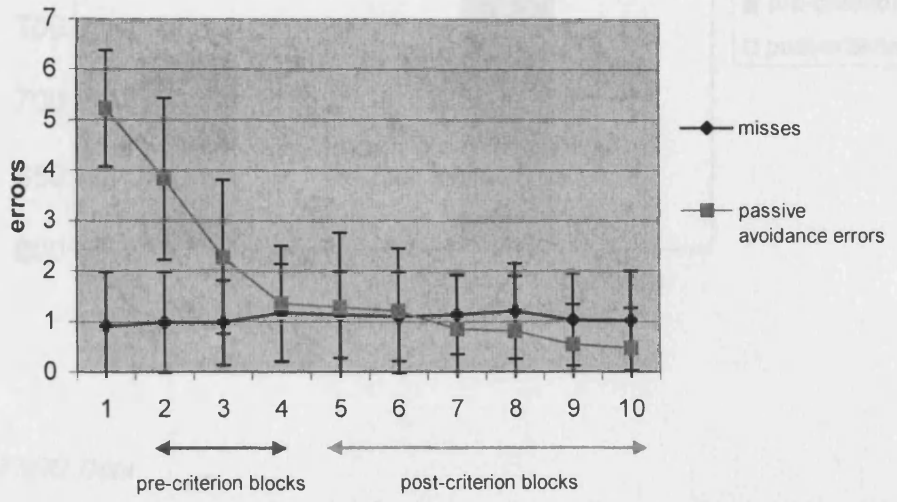
	Positive		Negative	
	Pre-criterion	Post-criterion	Pre-criterion	Post-criterion
Correct	39.32 (1.87)	76.84 (4.11)	27.00 (1.93)	79.31 (5.19)
Incorrect	8.05 (1.06)	17.90 (1.94)	20.37 (1.72)	15.42 (3.25)

(standard deviations in parentheses)

As can be seen in figure *3.2.*, participants showed clear indications of learning over blocks 1-4. From block 5 onwards, participants were regarded as task proficient. Following previous work (Blair et al., 2004; Newman and Kosson, 1986), data from block 1 were excluded. Consequently, trials in blocks 2-4 were labelled pre-criterion and trials in blocks 5-10 were labelled post-criterion. If the mean post-criterion passive avoidance error or correct rejection rate for a participant was greater than two standard deviations from the group

mean in a particular run, data for that run were excluded. This led to the exclusion of data from one run each for 7 of the 19 participants.

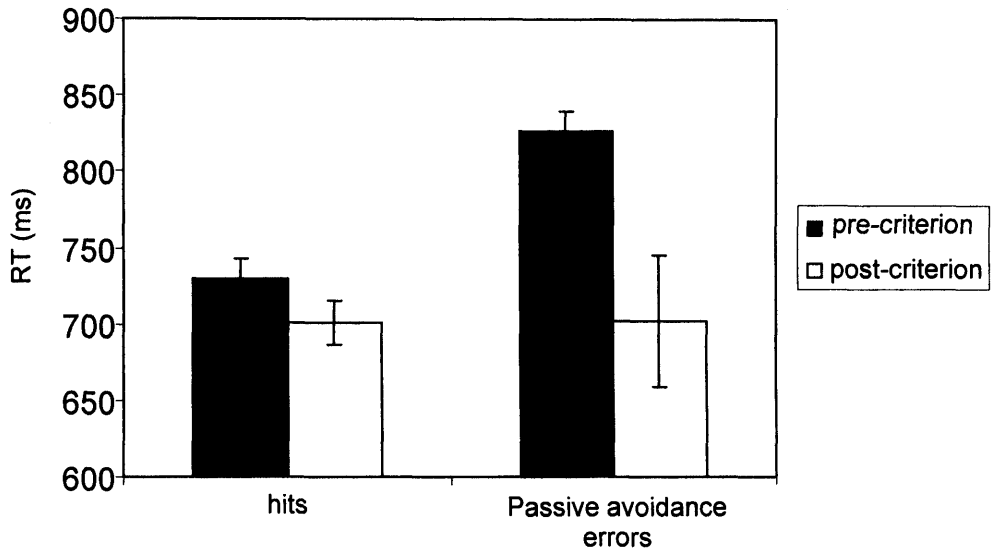
figure 3.2: Mean rates of passive avoidance errors and misses averaged across separate runs



(standard error bars) Red arrows indicate trials designated pre-criterion (blocks 2-4) and green arrows indicate trials designated post-criterion (blocks 5-10)

With respect to the participant's reaction times (RTs), these were significantly shorter for post-criterion hits relative to pre-criterion hits ($F(1, 18) = 9.61$; $P < 0.01$) (see figure *3.3.*). In addition, they were significantly shorter for post-criterion passive avoidance errors relative to pre-criterion passive avoidance errors ($F(1, 18) = 10.35$; $P < 0.005$).

figure. 3.3: RTs (ms) for correct and passive avoidance errors according to Learning Stage



3.3.2: FMRI Data

Post-criterion correct responses vs. all incorrect responses

The first contrast examined which regions showed significantly greater activation during performance of post-criterion correct responses relative to incorrect responses. This identified greater bilateral activation of anterior cingulate (ACC), left middle frontal gyrus, right posterior cingulate, left parietal lobe, bilateral caudate and left parahippocampal gyrus. This contrast was also performed on the anatomically defined ROI of the amygdala and revealed significant left amygdala activation (see table *3.2.* and figure *3.4.*).

Functionally defined ROIs were identified in left ACC, right posterior cingulate, left caudate and left parahippocampal gyrus. Mean BOLD responses in these ROIs were examined using a series of 2 (Stimulus Valence: 'good'/'bad') x 2 (Response Accuracy: correct/incorrect) x 2 (Learning Stage: pre-/ post-criterion) ANOVAs (table *3.2.* and figure *3.4.*). As expected, in each case, the main effect of Response Accuracy was significant. Left ACC ($P < 0.05$), right posterior cingulate ($P < 0.001$), left caudate ($P < 0.001$), left parahippocampal

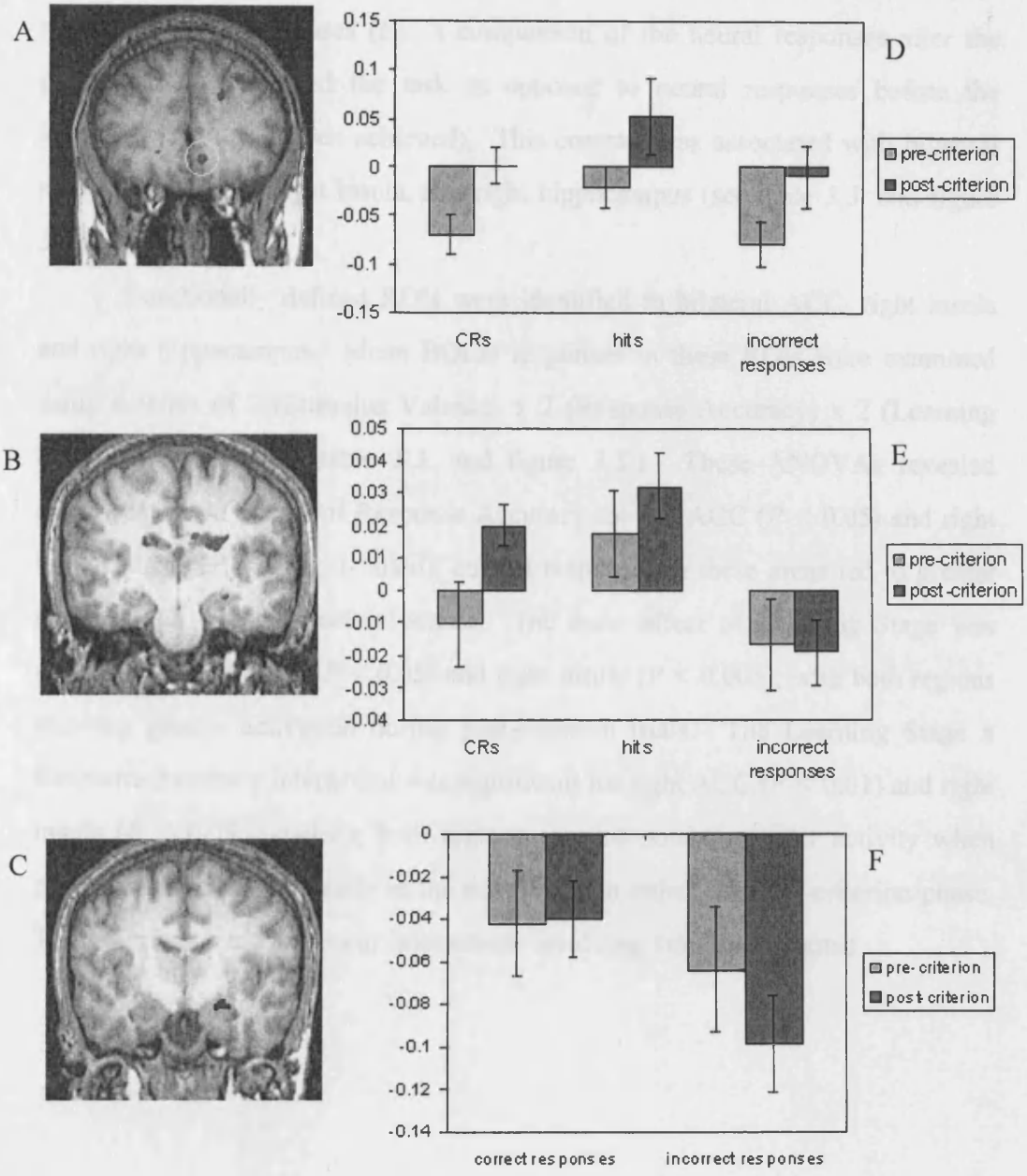
gyrus ($P < 0.01$), and left amygdala ($P < 0.005$) all yielded greater activation for correct than incorrect responses. The main effect of Learning Stage was significant in left ACC and right posterior cingulate cortex (both $P < 0.05$), with both regions showing greater responses during post-learning criterion trials rather than pre-learning criterion trials. There were no interactions between Response Accuracy and Learning Stage. There were also no main effects or interactions involving Stimulus Valence.

table 3.2: Significant differential activation produced by contrasting post-criterion correct responses with incorrect responses made throughout the task†

Anatomical location	l/r	BA	x	y	z	volume (mm^3)
Anterior Cingulate	l	24	-8	29	-7	247*
	r		15	32	2	331
Middle Frontal Gyrus	l	8	-20	33	42	432
Posterior cingulate	r	31	18	-37	30	1905
Inferior Parietal Lobule	l	39	-45	-70	43	599
Caudate & Caudate Body	l		-15	-19	25	3089
	r		14	-13	25	604
	r		21	-31	22	1034
Parahippocampal Gyrus	l		-24	-20	-10	117*
Amygdala	l		-16	-8	-16	73**

†All activations are effects observed in whole brain analyses corrected for multiple comparisons (significant at $P < 0.05$), except * significant at $P < 0.001$ uncorrected for multiple comparisons; and ** significant using an ROI analysis thresholded at $P < 0.05$.

figure 3.4: (A-F) Areas activated by post-criterion correct responses relative to incorrect responses



Activations are shown for: A left ACC; B left caudate; and, C left amygdala. The associated percentage signal change from baseline for these regions are shown in: D left ACC; E left caudate; and, F left amygdala.

Key: CR = correct rejection.

Post-criterion correct responses vs. pre-criterion correct responses

The second contrast examined which regions showed significantly greater activation during performance of post-criterion correct responses relative to pre-criterion correct responses (*i.e.*, a comparison of the neural responses after the participants had learned the task as opposed to neural responses before the learning plateau had been achieved). This contrast was associated with bilateral activation of ACC, right insula, and right hippocampus (see table *3.3.* and figure *3.5.*).

Functionally defined ROIs were identified in bilateral ACC, right insula and right hippocampus. Mean BOLD responses in these ROIs were examined using a series of 2 (Stimulus Valence) x 2 (Response Accuracy) x 2 (Learning Stage) ANOVAs (see table *3.3.* and figure *3.5.*). These ANOVAs revealed significant main effects of Response Accuracy for left ACC ($P < 0.05$) and right hippocampus ($P < 0.05$, 1-tailed); correct responses in these areas led to greater activity than did incorrect responses. The main effect of Learning Stage was significant in left ACC ($P < 0.05$) and right insula ($P < 0.005$), with both regions showing greater activation during post-criterion trials. The Learning Stage x Response Accuracy interaction was significant for right ACC ($P < 0.01$) and right insula ($P < 0.05$, 1-tailed); both regions showed notably greater activity when performing the task correctly in the post-criterion rather than pre-criterion phase. There were no main effects or interactions involving Stimulus Valence.

table 3.3: Significant differential activation produced by contrasting post-criterion correct responses with pre-criterion correct responses†

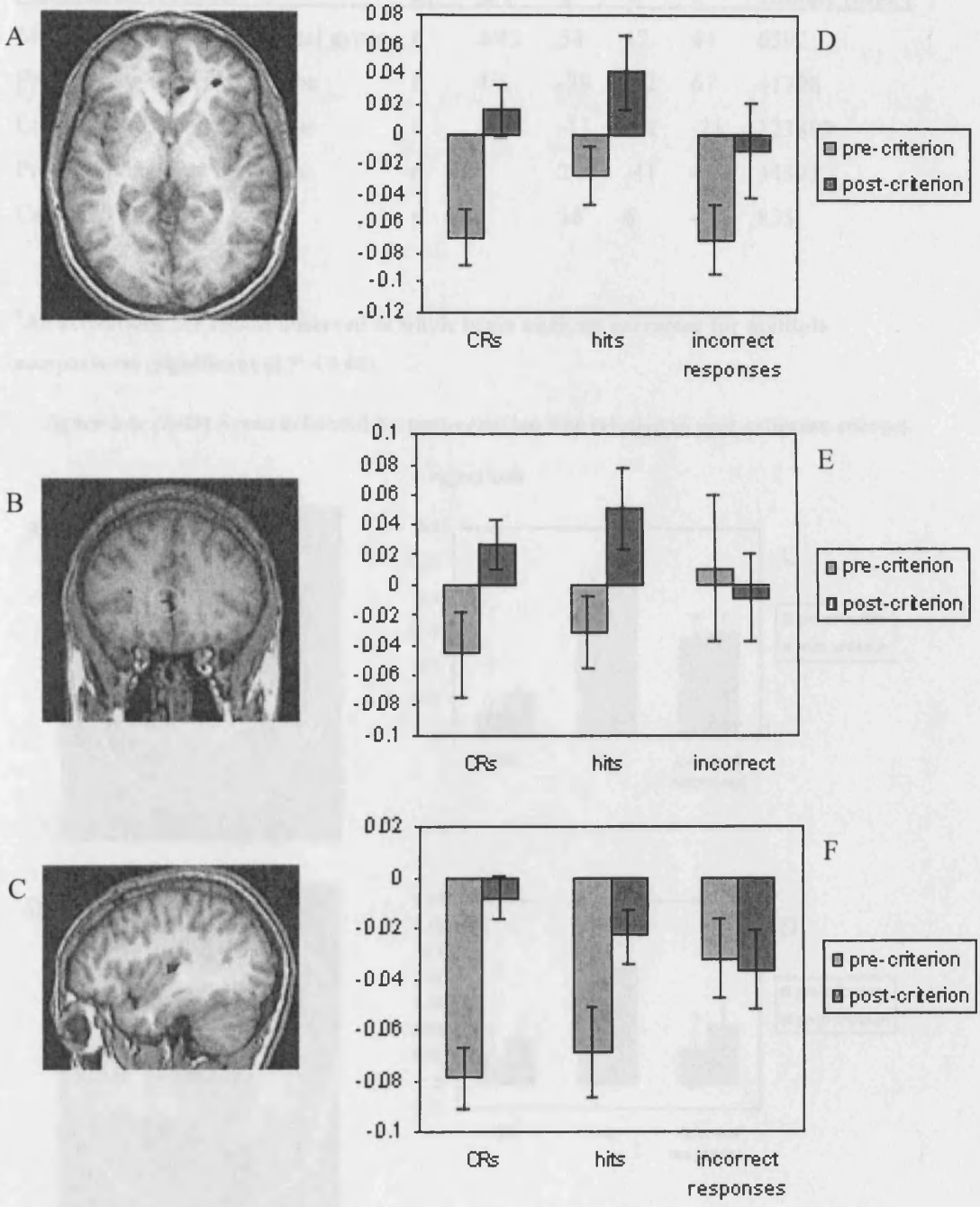
Anatomical location	l/r	BA	x	y	z	volume (mm^3)
Anterior Cingulate	l	24	-11	34	5	156*
	r		2	23	12	138*
	r		20	31	16	328
Insula	r		35	-15	11	944
Parahippocampal Gyrus	r		25	-16	-16	107*

†Activations are effects observed in whole brain analyses corrected for multiple comparisons (significant at $P < 0.05$), except * significant at $P < 0.001$ uncorrected for multiple comparisons.

Post-criterion hits vs. post-criterion correct rejections

The third contrast examined which regions showed significantly greater activation during performance of post-criterion hits relative to post-criterion correct rejections (*i.e.*, a comparison of the neural responses to 'good' vs. 'bad' stimuli after the participants had learned the task). This contrast revealed significantly greater activation within right middle frontal gyrus, left precentral gyrus, right posterior cingulate and right striatum (see table *3.4.* and figure *3.6.*). No regions showed greater activation to post-criterion correct rejections relative to post-criterion hits.

figure 3.5: (A-F) Areas activated by post-criterion correct responses relative to pre-criterion correct responses



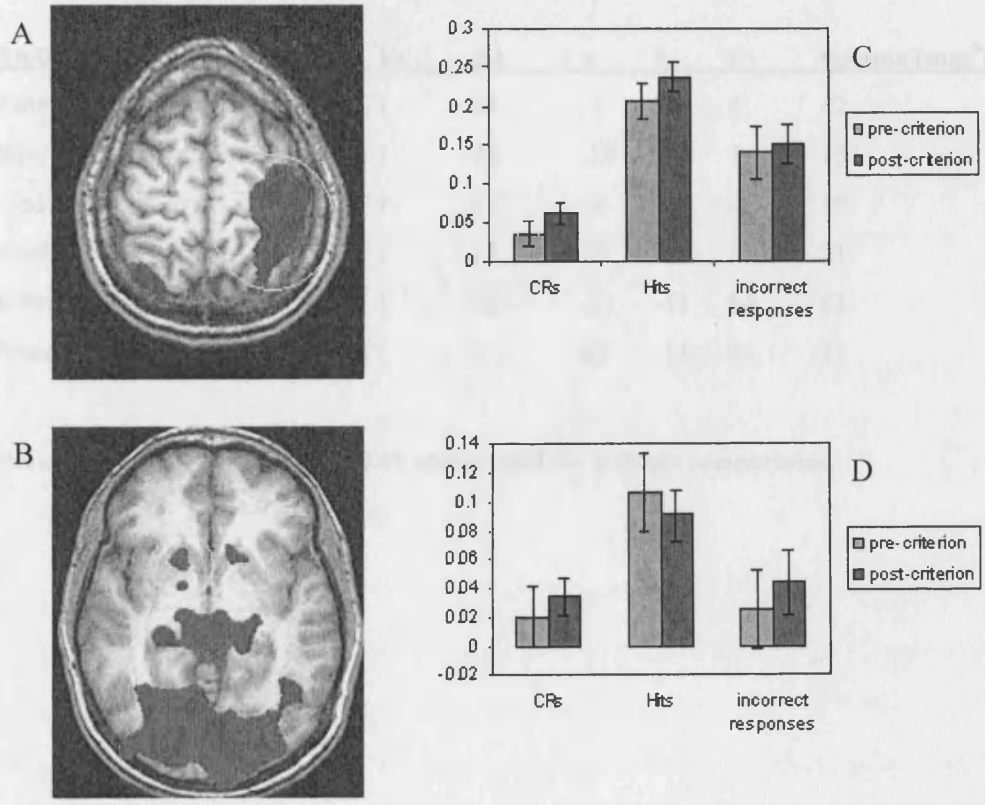
Activations are shown for: A left ACC; B right ACC; C right insula. The associated percentage signal changes from baseline for these regions are shown in: D left ACC; E right ACC; and, F right insula.
Key: CR = correct rejection.

table 3.4: Significant differential activation produced by contrasting post-criterion hits with post-criterion correct rejections †

Anatomical location	l/r	BA	x	y	z	volume (mm^3)
Middle frontal/inferior frontal gyrus	r	8/45	58	12	44	6592
Precentral gyrus/Parietal lobe	l	4/6	-39	-22	67	41798
Lingual gyrus/fusiform gyrus	l	37	-11	-88	-21	123809
Posterior cingulate/thalamus	r		2	-41	44	34393
Caudate/Putamen	r		16	6	-7	855

†All activations are effects observed in whole brain analyses corrected for multiple comparisons (significant at $P < 0.05$).

figure 3.6: (A-D) Areas activated by post-criterion hits relative to post-criterion correct rejections



Activations are shown for: A left precentral gyrus; and, B right caudate. The associated percentage signal changes from baseline for these regions are shown in: C left precentral gyrus; and D right caudate.

Key: CR = correct rejection.

Conjunction analysis (post-criterion hits vs. pre-criterion hits plus post-criterion correct rejections vs. pre-criterion correct rejections)

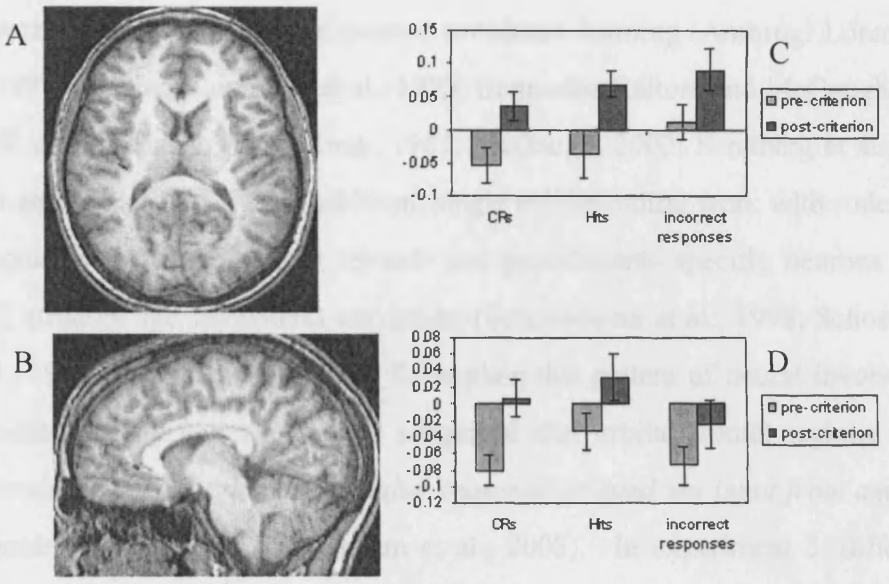
The conjunction analysis examined activity common to post-criterion hits vs. pre-criterion hits and post-criterion correct rejections vs. pre-criterion correct rejections. Increased learning related activation common to both 'good' and 'bad' stimuli was seen in both ACC and insula (see table *3.5.* and figure *3.7.*).

table 3.5: Significant differential activation produced in a conjunction of post-criterion hits vs. pre-criterion hits and post-criterion correct rejections vs. pre-criterion correct rejections †

Anatomical location	l/r	BA	x	y	z	volume (mm^3)
Anterior Cingulate	l	24	-7	33	4	42
Inferior frontal gyrus	l	45	-29	35	7	151
Medial frontal gyrus	r	11	10	53	-17	19
Insula	r	13	35	-20	13	21
Inferior parietal lobe	l	40	-57	-41	54	13
Precentral gyrus	r	6	63	-14	44	21

†All activations significant at $P < 0.01$ uncorrected for multiple comparisons.

figure 3.7: (A-D) Areas activated in the conjunction of post-criterion hits (relative to pre-criterion hits) and post-criterion correct rejections (relative to pre-criterion correct rejections)



Activations are shown for: A left insula; and, B left ACC. The associated percentage signal changes from baseline for these regions are shown in: C left insula; and, D left ACC. Key: CR = correct rejection.

3.4: Discussion

Experiment 3 investigated the neural substrates involved during successful performance of a task of passive avoidance learning. Specifically, the pattern of activation associated with correct responses performed when participants were deemed to be task-proficient was contrasted with those elicited by incorrect responses and also correct responses performed prior to task-proficiency. The contrasts revealed that correct responses following successful learning of the task were associated with activation in regions including rostral ACC, insula, caudate, hippocampal regions, and the amygdala. Behaviourally, the participants showed a distinct pattern of performance, with initial exploratory responding (pre-criterion

performance) followed by relatively rapid learning before a stable plateau of successful responding (post-criterion performance) was achieved (see figure *3.4.*).

The results obtained in experiment 3 were broadly in line with the animal literature. Thus, animal work has indicated the importance of the insula, striatum, hippocampus and amygdala in passive avoidance learning (Ambrogi Lorenzini et al., 1997; Ambrogi Lorenzini et al., 1995; Bermudez-Rattoni and McGaugh, 1991; Cahill and McGaugh, 1990; Gray, 1987; McGaugh, 2002; Sandberg et al., 1984; Treit and Menard, 1997). In addition, single cell recording work with rodents has demonstrated that there exist reward- and punishment- specific neurons within OFC, striatum and basolateral amygdala, (Schoenbaum et al., 1998; Schoenbaum et al., 1999; Setlow et al., 2003). To explain this pattern of neural involvement, Schoenbaum and colleagues have suggested that orbital frontal regions *"apply information regarding incentive value acquired or cued via input from amygdala to guide responding"* (Schoenbaum et al., 2003). In experiment 3, differential activation in rostral ACC, insula, caudate, hippocampal regions, and amygdala was observed to post-criterion correct responses relative to incorrect responses and pre-criterion correct responses (see figures. *3.4.* and *3.5.*). It must be noted, however, that in the case of the insula and amygdala the differential activation reflected decreased deactivation making interpretation difficult. Even so, it appears that the pattern of activation here may be considered consistent with the idea that both the amygdala and insula provide input regarding the incentive value of the current stimulus to rostral ACC, which, in turn, guides behavioural responses which are mediated by the caudate.

Whilst the primary frontal region identified in experiment 3 was rostral ACC, it must be noted that much of the single cell recording work with animals has focused upon a posterior region of medial OFC (Schoenbaum et al., 1998; Schoenbaum et al., 1999; Setlow et al., 2003). It must be noted however that the rostral ACC activation observed here is within an area of ACC which has been considered the 'emotional' region (Bush et al., 2000; Rogers et al, 2004b; Whalen et al., 1998). This area of ACC projects densely to OFC, striatum, hypothalamus, and brain stem; all areas that have been implicated in emotional learning.

Moreover, recent studies have reported activation in this region in response to positive outcomes during decision making (Rogers et al., 2004b) and also during amphetamine administration (Vollm et al., 2004). Additionally, as noted in section *3.1.* Elliott et al (Elliott et al., 2000b) observed reward-related activity in subgenual cingulate during a decision making paradigm, and Cox et al (Cox et al., 2005) reported medial OFC/ rostral ACC activity both with respect to reward itself as well as to conditioned stimuli that predicted reward. Furthermore, activations of similar medial regions have been seen in response to reward in the context of reversal learning paradigms (O'Doherty et al., 2001; O'Doherty et al., 2003b). Several researchers have suggested that this medial region of frontal cortex is particularly involved in processing rewarding stimuli (Elliott et al., 2000a; O'Doherty et al., 2001; Rolls, 2000). Alternatively, medial frontal cortex/ rostral ACC has been considered to serve a role in decision making/ response selection (Bechara et al., 2000a), perhaps utilizing expected reinforcement information to guide response selection and stimulus choice (Blair, 2004). Under these circumstances, medial frontal cortex/ rostral ACC activation might be anticipated to reflect expectancies of incentive value. This explanation may account for the activation in rostral ACC in this experiment, and is in line with data from the animal literature (Schoenbaum et al., 1998; Schoenbaum et al., 1999; Setlow et al., 2003).

Importantly, the increased activation in rostral ACC, insula, caudate, hippocampal regions and amygdala to post-criterion correct responses relative to incorrect responses or pre-criterion correct responses was observed to both CS+s and CS-s, *i.e.*, when approaching 'good' stimuli *and* avoiding 'bad' stimuli (see figures. *3.4.* and *3.5.*). This was also seen with respect to rostral ACC and insula in the conjunction analysis (see figure *3.7.*). This is consistent with the proposal that stimulus-reward and stimulus-punishment associations are both important influences on behaviour and choice (Baxter and Murray, 2002). In addition, it is consistent with recent data suggesting that the amygdala, striatum and medial frontal regions are involved in processing both appetitive and aversive stimuli (Baxter and Murray, 2002; Everitt et al., 2003; Seymour et al., 2004). It is also

worth noting that the current results cannot be attributed to outcome information. Notably, the increased neural responses in rostral ACC, insula, caudate, hippocampal regions, and amygdala to post-criterion correct responses was seen when participants made correct rejections; *i.e.*, on trials when no performance-related feedback was given.

Previous work examining the response of the caudate to reinforcement information has frequently reported increased activity to reward information and decreased activity to punishment information (Delgado et al., 2003; Delgado et al., 2000; O'Doherty et al., 2003b; Rogers et al., 2000). In all of these paradigms participants responded to stimuli in the expectation of reward but sometimes received punishment. Consequently it was reported that the expectation and receipt of reward were associated with greater caudate activity than the expectation and receipt of punishment. This activation, however, may simply have reflected stronger stimulus-response associations following rewarded responses rather than punished responses. In the current experiment increased activation in caudate was observed to 'good' stimuli in comparison with 'bad' stimuli (see figures *3.4.* and *3.6.*). However, it is worth noting that there was also increased activation in caudate to correct rejections following successful learning relative to correct rejections before successful learning had been achieved (see figure *3.4.*). In other words, there was increased caudate activity to the 'bad' stimuli as a function of learning. Importantly, in the current task, following learning, participants avoided the 'bad' stimuli in order to avoid punishment. In contrast with other neuroimaging tasks participants were not 'surprised' by an unexpected punishment [*e.g.* the decision making task of Delgado and colleagues (Delgado et al., 2000) or the reversal learning paradigms used by O'Doherty and colleagues (O'Doherty et al., 2003b) and Rogers and colleagues (Rogers et al., 2000)]. It may be suggested that the increased caudate activity seen in experiment 3 was a partial function of reinforcement outcome expectancy information. In the case of the 'bad' stimuli, either an expectancy of the punishment that would be avoided by withholding a response or even that the act of avoiding a punishment was in itself a rewarding experience.

Finally, it is worth comparing the results of the third contrast (post-criterion hits vs. post-criterion correct rejections) to the results of previous studies examining the neural systems mediating performance on the human go/no-go task (*e.g.* Casey et al., 2001; Konishi et al., 1998; Liddle et al., 2001). In these tasks, participants are instructed to respond to one class of stimuli but refrain from responding to another. For example, in the study by Casey and colleagues (Casey et al., 2001), participants were instructed to respond to any letter other than 'X'. A key difference between this go/no-go task and the passive avoidance learning task presented in experiment 3 is that it does not require the formation of stimulus-reinforcement representations. Instead participants are explicitly informed which stimuli should be approached and which should be avoided. Correct 'go' responses in go/no-go tasks have been associated with greater activity in the supplementary motor area (*e.g.* Liddle et al., 2001; Watanabe et al., 2002). Increased BOLD activation within this region, as well as caudate, was observed in experiment 3 to post-criterion hits. This activation probably reflected the recruitment of neural systems to mediate the motor response. In contrast, correct 'no-go' responses in go/no-go tasks have been associated with greater activity in both dorsolateral and ventrolateral PFC. It has been commonly suggested that these regions are recruited in order to inhibit a motor response (*e.g.*, Casey et al., 2001; Konishi et al., 1998; Liddle et al., 2001; Watanabe et al., 2002). In terms of the passive avoidance learning task the avoidance response is hypothesized to be activated, as a function of outcome expectancy information (see sections *1.3.6.* and *2.1.*), that is, the act of avoidance is a response, over and above inhibition of an approach response. In line with this suggestion, the converse contrast (post-criterion correct rejections vs. post-criterion hits) was not associated with significant differential activation in any region, including ventrolateral PFC.

3.4.1: Summary and Conclusions

In conclusion, experiment 3 investigated the neural substrates involved in successful performance on a passive avoidance learning task. It was revealed,

consistent with the animal literature, that successful passive avoidance learning requires the appropriate recruitment of regions including rostral ACC, insula, caudate, hippocampal regions, and the amygdala.

3.5: General Discussion

The results of experiment 3 strengthen claims, in line with the integrated emotion systems (IES) model, that stimulus-reinforcement learning recruits a circuit including amygdala and medial frontal regions. Further, it may also be inferred then that experiment 3 lends further support that dysfunction within this circuit may be present in individuals with psychopathy.

The results of experiment 3 were largely in line with predictions of the IES model. Specifically, results indicated that stimulus-reward and stimulus-punishment associations were mediated by the same circuit – a prediction of the IES account (Blair, 2004). Moreover, these data were in accordance with the conceptualisation that successful performance of passive avoidance learning reflects acquisition of valence representations. Indeed, the results obtained here were dissimilar to those that would have been expected had there been no emotional learning involved (*i.e.* in 'go/no-go' tasks). In contrast with IES model, however, the data from experiment 3 appear to indicate that rostral ACC/medial OFC, rather than ventrolateral PFC, is important in the detection of expectation violations. It is also interesting that the insula was activated in this task, indeed, the IES model suggests that the insula may be a locus for the storage of affect representations that are formed within the amygdala. This hypothesis was asserted based on the observation that adults sustaining amygdala lesions do not 'lose' previously acquired affect representations (Blair, 2004). These data, therefore, indicate that whilst the predictions of the IES model are relatively accurate, it requires some modification if it is to be entirely successful in explaining passive avoidance learning.

3.6: Conclusions

In conclusion, experiment 3 investigated the neural substrates involved in successful passive avoidance learning. In line with predictions, based on the IES account, correct responses to both rewarding and punishing stimuli (*i.e.* approaches and avoidances respectively) recruited similar regions, including amygdala, rostral ACC, insula, caudate and hippocampus.

Chapter 4 – Probabilistic Reversal Learning in Children with Psychopathic Tendencies and Adult Psychopaths

4.1: Introduction to Reversal Learning

Reversal learning is the alteration of a behavioural response to a previously rewarded stimulus in accordance with reversed reinforcement contingencies (Rolls, 1999). It is a cognitive flexibility that enables individuals to engage efficiently in the pursuit of rewards and the avoidance of punishments, and is thus important in a wide range of motivated and emotional behaviours (Rolls, 1999). Reversal learning tasks usually involve separate acquisition and reversal phases. In the acquisition phase participants learn an object discrimination in order to gain reward (*i.e.*, '*if presented with stimulus pair A-B, choose A*'). Following satisfactory performance reinforcement contingencies are reversed, and in order to continue receiving reward (and avoiding punishment) participants are required to reverse their stimulus choice (*i.e.*, '*if presented with stimulus pair A-B, choose B*'). Importantly, and similarly to passive avoidance learning, participants are not informed of the reward- and punishment-contingencies prior to task performance. Instead they must learn by trial-and-error which is the correct stimulus to select.

Extinction is a cognitive function that is conceptually similar to reversal learning. Instead of contingency reversal, during the second phase of the task a 'no response' rule becomes correct (*i.e.* '*if presented with stimulus A, withhold response*'). Thus in order to avoid punishment (and sometimes to receive reward) participants are required to cease (*i.e.* extinguish) responding.

Reversal and extinction tasks usually involve presentation of stimuli in a pair-wise fashion over the course of the experiment (*e.g.* Cools et al., 2002; Downes et al., 1989; Kringelbach and Rolls, 2003; O'Doherty et al., 2003a; O'Doherty et al., 2001; Remijnse et al., 2005). According to Baxter and Murray (2002) this type of instrumental learning, in contrast to passive avoidance learning, may be solved by forming stimulus-response associations (see section

1.3.6.). That is, by forming an association between a conditioned stimulus (CS) and a motor response. It has been suggested that following repeated pair-wise presentations, two stimuli (*i.e.* A and B) come to form a single compound representation within temporal cortex (*i.e.* stimulus A-B) (Messinger et al., 2001). As such, this compound stimulus A-B may be either good or bad, dependent upon the behaviour of the participant (*i.e.* whether they do action X or Y). It is thought that this type of stimulus-response learning may be easily solved by means of an interaction between temporal cortex and caudate, which commands a particular response (*e.g.* perform action X) when the compound stimulus A-B is presented (Packard, 1999; Packard and McGaugh, 1996).

Commonly, reversal learning ability has been assessed by counting the number of errors made before the participant meets a criterion of consecutive correct responses. Alternatively performance may be assessed by noting the proportion of adaptive versus maladaptive responses (Berlin et al., 2004). In this type of 'win-stay, lose-shift' analysis responses are defined according to feedback received on the previous trial (*i.e.* 'win' or 'lose') in combination with the action committed on the current trial (*i.e.* 'stay' [reapproach the stimulus approached on the previous trial] or 'shift' [approach the alternative stimulus]). In simple reversal learning paradigms, adaptive responses are to win-stay and to lose-shift (see figure *4.1.*).

figure 4.1: A representation of all possible responses in a win-stay, lose-shift analysis of simple reversal learning

		Action	
		Stay	*Shift*
Previous	*Win*	win-stay	win-shift
Feedback	*Lose*	Lose-stay	lose-shift

Adaptive responses highlighted in green and maladaptive responses highlighted in red

Recently researchers have employed probabilistic reversal learning tasks (*e.g.* Lawrence et al., 1999; O'Doherty et al., 2001; Swainson et al., 2000). In these tasks the reinforcement contingencies are inconsistent. For example, the 'correct' stimulus in a pair with an 80-20 probabilistic reinforcement contingency would be rewarded on 8 out of 10 presentations and punished on 2 out of 10 presentations, and the opposite reinforcement contingency would be true of the 'incorrect' stimulus (Lawrence et al., 1999; Swainson et al., 2000). Probabilistic contingencies have been used to simulate 'real life' problems where there is often less than a 1:1 matching between a stimulus and a reward (Lawrence et al., 1999). Importantly, sometimes patients present with selective impairments in performance of probabilistic tasks (*e.g.* Lawrence et al., 1999). In probabilistic tasks a consistent win-stay, lose-shift strategy becomes maladaptive. Instead participants must learn to lose-stay on some trials, which may be termed *'probabilistic error'* trials (Cools et al., 2002; Swainson et al., 2000).

4.2: Experiment 4

As introduced in chapter 1, adult individuals with psychopathy have consistently demonstrated impaired reversal learning and extinction (Mitchell et al., 2002; Newman et al., 1987). The case is less clear, however, as regards children with psychopathic tendencies (Blair et al., 2001a; Fisher and Blair, 1998; O'Brien and Frick, 1996). Experiment 4 aims to investigate and further characterise the reversal learning impairment demonstrated by children with psychopathic tendencies.

Bechara's Iowa gambling task is a complex reversal learning paradigm (Bechara et al., 1994; Bechara et al., 1999; Bechara et al., 2000b). On each trial participants must sample from one of four decks of cards. Reward or punishment (which takes the form of award or removal of hypothetical money) is issued dependent upon probabilistically-defined reinforcement contingencies. Importantly, *initially selection from all decks leads to reward.* However, as the task progresses, consistent selection from two of the decks – the risky decks – leads to a net loss (due to the high rewards, and higher punishments associated with these decks). Inversely, consistent selection from the other two decks – the safe decks – leads to a net gain (due to the low rewards, and lower punishments). Participants usually begin the task by selecting cards from the risky decks, in order to redeem the associated high rewards. The imperative of the task (to win as much money as possible), however, may be achieved only by reversing responses away from the risky decks towards the safe decks. Indeed, healthy participants learn relatively quickly to reverse their responses towards the safe decks (Bechara et al., 1994; Bechara et al., 1997). In contrast, performance of individuals with psychopathy is characterized by consistent selection from risky decks. Deficits on this task have been observed in both adult psychopaths and children with psychopathic tendencies (Blair et al., 2001a; Mitchell et al., 2002). Although a complex task, the encouragement of an initial preference towards the risky decks, coupled with the introduction of increased punishment associated with these decks

has led to suggestions that this task essentially assesses reversal learning ability (Blair et al., 2001a; Fellows and Farah, 2005; Mitchell et al., 2002).

Individuals with psychopathy have also presented with impairment on Newman's card task, which may be described as an extinction task under partial reinforcement conditions (Newman et al., 1987). In this task participants must decide when to discontinue playing cards, from a single deck of 100, in order to maintain maximum points. Whilst initially all responses are rewarded (with 5 cents), as the task proceeds the reward to punishment ratio decreases in 10% increments for each 10 cards played. This leads to a higher punishment rate as the task progresses. Adult psychopaths and children with psychopathic tendencies have consistently played significantly more cards than controls, leading to a lower cumulative reward (Fisher and Blair, 1998; Newman et al., 1987; O'Brien and Frick, 1996).

Finally, the IDED task is a multistage instrumental learning task that incorporates a simple reversal condition (Downes et al., 1989). In short, two stimuli are presented concurrently on a computer screen and after a pre-specified number of trials as the correct stimulus, the contingencies reverse and the previously correct stimulus suddenly becomes incorrect (and *vice versa*). In order to continue receiving rewarding feedback participants must reverse their responses away from the previously correct (now incorrect) stimulus, towards the newly correct stimulus. Whilst adults with psychopathy have presented with impairment on the task (Mitchell et al., 2002), in contrast, children with psychopathic tendencies have not (Blair et al., 2001a).

A commonality between the Iowa gambling task and Newman's extinction task is that they involve complex reinforcement contingencies. Due to the design of these tasks the contingency changes are subtle. In contrast, the IDED task, on which children with psychopathic tendencies perform successfully, involves a simple reinforcement contingency. In this task, upon reversal, the previously *consistently* rewarded stimulus is now *consistently* punished. The contingency change, therefore, is more easily perceptible in the IDED task than in either the Iowa gambling task or Newman's extinction task. It thus appears that the salience

of the contingency change may be a critical factor determining the success of children with psychopathic tendencies on these tasks. In short, it appears that they may be sensitive to highly salient contingency changes, and less sensitive to more subtle contingency changes.

A further commonality between the two tasks on which children with psychopathic tendencies perform unsuccessfully are their complex natures. The Iowa task in particular involves other components unrelated to reversal ability, such as the ability to attend to, synthesize and remember complex reinforcement histories and to resolve the approach-avoidance conflicts that arise when a deck is associated with both reward and punishment (Busemeyer and Stout, 2002; Fellows and Farah, 2005; Rogers et al., 1999b). Also, Newman's task assesses extinction and not reversal (Newman et al., 1987). Whilst these cognitive functions are conceptually related, they are not directly equivalent.

As is the case with passive avoidance learning, a recent study has indicated that children with attention-deficit/hyperactivity disorder (ADHD) may also present with impaired reversal learning performance. Itami and Uno (2002) tested children with ADHD (predominantly hyperactive-impulsive, or combined type) on tasks assessing reversal and extinction. Results indicated that the ADHD group committed more errors, on both tasks, than comparison children. Due to the high co-morbidity between ADHD and psychopathic tendencies it is important to clarify which disorder to driving the impairment (Colledge and Blair, 2001).

4.2.1: Summary of Aims

Experiment 4 attempted to assess the performance of children with psychopathic tendencies on a novel probabilistic reversal learning paradigm with four different contingencies (100-0; 90-10; 80-20; 70-30). Specifically experiment 4 aimed to determine the extent of the reversal learning impairment present, and in particular to determine whether any group differences were significantly related to psychopathic tendencies after the variation due to level of ADHD has been removed.

4.3: Methods

4.3.1: Design

The independent variables were: group (psychopathic tendencies/ comparisons); the two phases (acquisition of the discrimination/reversal of the discrimination); and the four reinforcement contingencies (100-0/90-10/80-20/70-30). The dependent variable was errors to criterion (see section *4.3.4.* for details).

4.3.2: Participants

The participants were all boys aged between 8 and 16 years recruited from three UK government-run schools for children with emotional and behavioural difficulties. They had all received statements under the Education Act of 1993 as being too problematic for mainstream education. Of the boys taking part, 38 were of Caucasian origin, the remaining 3 were of Afro-Caribbean origin (one of whom was in the psychopathic tendencies group).

Initially all boys whose parents' granted consent were screened using the Antisocial Process Screening Device (Frick & Hare, 2001). The participants were selected on the basis of the combined APSD scores of two raters (usually two teachers but on occasion, a teacher and a classroom assistant). In line with previous work (Blair et al., 2001c; Fisher and Blair, 1998), participants with an APSD score of 27 or above were eligible for the psychopathic tendencies group and participants with an APSD score of 15 or below were eligible for the comparison group. Eighteen boys were included in the psychopathic tendencies group and 22 boys were included in the comparison group[3]. It was made clear to the participants that they were free to withdraw from the study at any time.

[3] 8 children with psychopathic tendencies, and 5 comparison children from experiment 1 also participated in this experiment.

4.3.3: Measures

British Picture Vocabulary Scale (BPVS; Dunn, Wheklan & Pintillie, 1982).

The BPVS was used to measure the participants' verbal intelligence quotient (IQ). The BPVS measures receptive vocabulary for standard English.

Antisocial Process Screening Device (APSD; Frick & Hare, 2001)

Participant's scores for each item were the averages assigned by the two raters. Pearson's correlations of the two ratings for each child were $r2 = 0.80$ ($P < 0.001$) for total APSD score. Inter-rater correlations for the three factors were: $r2 = 0.43$ ($P < 0.01$) for callous/unemotional; $r2 = 0.82$ ($P < 0.001$) for narcissism, and; $r2 = 0.68$ ($P < 0.001$) for impulsivity.

ADHD Rating Scale - IV (DuPaul et al., 1998)

Participant's scores for each item were the averages assigned by the two raters. The Pearson's correlation of the ratings was $r2 = 0.63$ ($P < 0.001$) for total DuPaul. Inter-rater correlations for the two factors were $r2 = 0.59$ ($P < 0.001$) for hyperactivity-impulsivity and $r2 = 0.63$ ($P < 0.001$) for inattention.

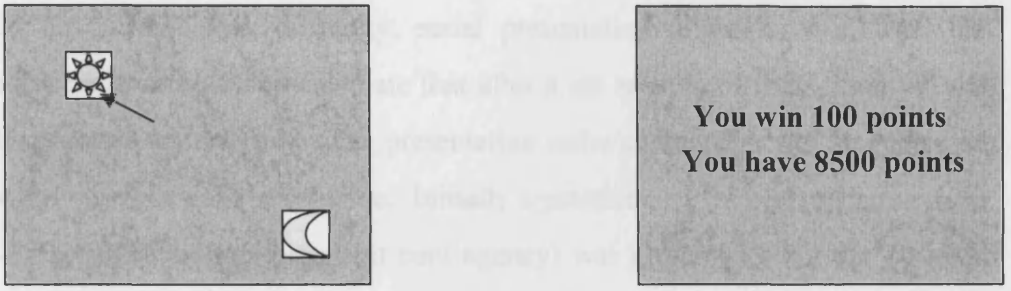
4.3.4: Probabilistic Reversal Learning Task.

The task was programmed in Visual Basic (6.0) and was presented on a Dell Laptop computer. Stimuli comprised 16 line drawings of animals (Snodgrass and Vanderwart, 1980) which were shaded in different colours (see figure *4.2.*). Stimuli measured 4 cm by 4 cm and were presented on a grey background.

Stimuli were assigned into pairs randomly at the beginning of the task. On each trial, a pair of stimuli was presented. Stimulus locations were assigned randomly on each trial (there were 16 possible locations). The participant had to choose one of the stimuli by clicking on it with the mouse. Upon choosing they would receive either positive ('you win 100 points') or negative ('you lose 100 points') feedback (for 1000 ms) depending on the reinforcement contingency of that pair. One of the animals in each pair was always more likely than the other to

be rewarded rather than punished. Participants began the task with 0 points. A running total of points was visible at the bottom of the screen only during the feedback display. Trials were self-paced.

figure 4.2: An example of a trial in the probabilistic reversal learning task
The participant approached the sun stimulus and was rewarded with 100 points



Self-paced selection

Feedback for 1000 ms

The reinforcement contingencies were probabilistic such that the 'correct' pair was not always rewarded and the 'incorrect' pair was not always punished. The 'correct' stimulus in each pair was always the one with the greater ratio of reward to punishment. For example the 'correct' stimulus in a pair with a 90-10 reward to punishment contingency would be rewarded on 9 out of every 10 trials and punished on 1 out of every 10 trials. Inversely, the 'incorrect' stimulus would be punished on 9 out of every 10 trials and rewarded on 1 out of every 10 trials. The order of probabilistic feedback was randomised within the program.

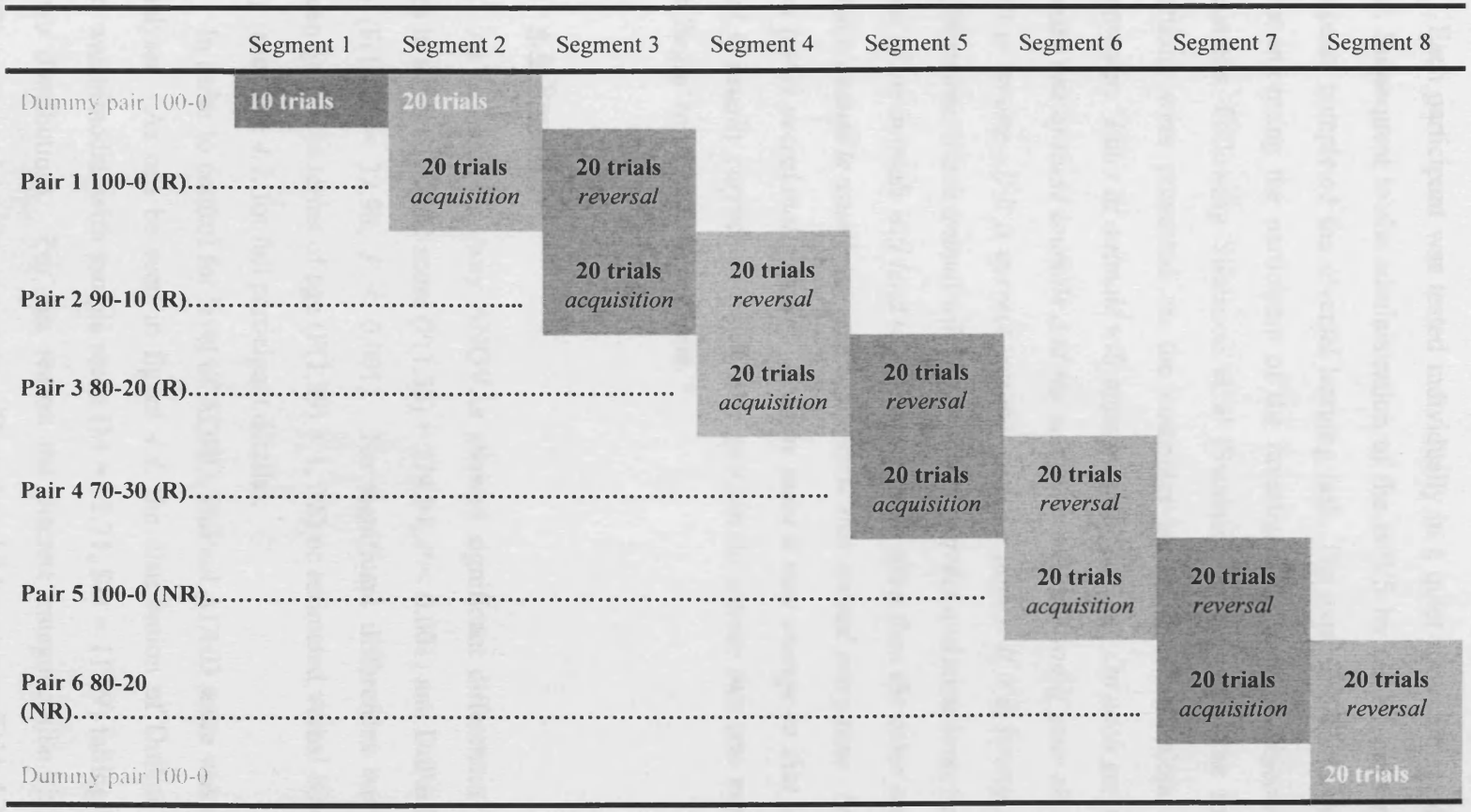
There were six different experimental pairs of stimuli: four that changed contingency (reversing pairs; R) and two that did not (non-reversing pairs; NR) (see figure *4.3.*). The four reversing pairs had the following probabilistic contingencies: 100-0; 90-10; 80-20; and 70-30). The reinforcement contingency of the four pairs remained constant for 20 trials (phase 1; acquisition of the discrimination). Upon completing 20 trials the reinforcement contingency was reversed (phase 2; reversal of the discrimination). Thus the previously correct stimulus became the incorrect stimulus and the previously incorrect stimulus now became the correct stimulus. The two non-reversing pairs had reinforcement

contingencies 90-10 and 70-30. The contingencies of these pairs remained constant for the 40 trials for which they were presented.

Rather than learn about pairs of stimuli serially (*e.g.*, all 40 trials of the 100-0 contingency followed by all 40 trials of 80-20 contingency, and so on), participants had concurrent experience with two different pairs of stimuli at all times (*i.e.* trials from two pairs were presented alternately). This was done in order to increase task difficulty; serial presentation might have allowed the participant to more easily calculate that after a set amount of trials, many of the pairs changed contingency. The presentation order of the pairs was staggered so that the reversals did not coincide. Initially a practice pair (non-reversing with an 80-20 probabilistic reinforcement contingency) was presented alone for 10 trials (to introduce participants to probabilistic reinforcement) and then alternately with the first experimental pair. Towards the end of the task a dummy pair was introduced and presented on alternate trials with the final experimental pair. Thus each experimental pair was presented on alternate trials with another pair of stimuli. All participants received 290 trials regardless of performance. The order of pair presentation was randomised.

For the purposes of analysis a learning criterion of 6 consecutive correct responses was imposed in both phases. Thus participants had to choose the correct stimulus in each pair six times consecutively before they had successfully passed that phase of the task. If participants did not meet the learning criterion in the acquisition phase, total errors made were analysed in the acquisition phase, and they were excluded from reversal phase analyses.

figure 4.3: A schematic representation of the probabilistic reversal learning task



The presentation order of pairs 1-6 was randomised. In each segment trials from pairs were presented alternately

Key to figure 1: R = reversing, NR = non-reversing

4.3.5: Procedure

Each participant was tested individually in a quiet room allocated by the school. Subsequent to the administration of the BPVS by the experimenter, the participants completed the reversal learning task. The experiment was described without informing the participant of the investigation's specific objectives and expectations. Following Swainson et al (Swainson et al., 2000) the following instructions were presented on the computer screen and read aloud by the experimenter: '*Pairs of animals will appear on the screen. On each go you have to choose one of these animals and the computer will tell you if your choice was correct or wrong. If it is correct you will win 100 points. If it is wrong you will lose 100 points. Each animal will sometimes be correct and sometimes be wrong, but one of the animals will tend to be correct more often than the other one. Find out which animal is usually correct, and choose that animal every time. Stick with it even if it is occasionally wrong. At some point it may change so that the other animal is usually correct, in which case you should choose that one every time. Press 'begin' to start the experiment.'*

4.4: Results

As expected one-way ANOVAs showed significant differences between groups in terms of APSD score ($F(1,38) = 374.94$, $P < 0.001$) and DuPaul ADHD score ($F(1,38) = 32.98$, $P < 0.001$). No significant differences were found between groups in terms of age ($F(1,39) < 1$, *NS*) or estimated verbal IQ ($F(1,39) <$ *NS*) (see table *4.1.* for full participant details).

In order to control for level of ADHD, DuPaul ADHD score was covaried in analyses. As can be seen in figure *4.4.* the distribution of DuPaul ADHD scores was bimodal, with sample mean ($M = 22.71$, $SD = 11.89$) falling between the two distributions. For this reason the discrete categorization (high/ low ADHD), was used as the covariate. Thus the children were divided into two groups (high/low ADHD) according to a median split of scores on the DuPaul ADHD rating scale; those children scoring up to and including 24 points were

included in the low ADHD group (n = 20), whilst those scoring above 24 were included in the high ADHD group (n = 20). Three of the boys presenting with psychopathic tendencies were included in the low ADHD group and five of the comparison boys were included in the high ADHD group.

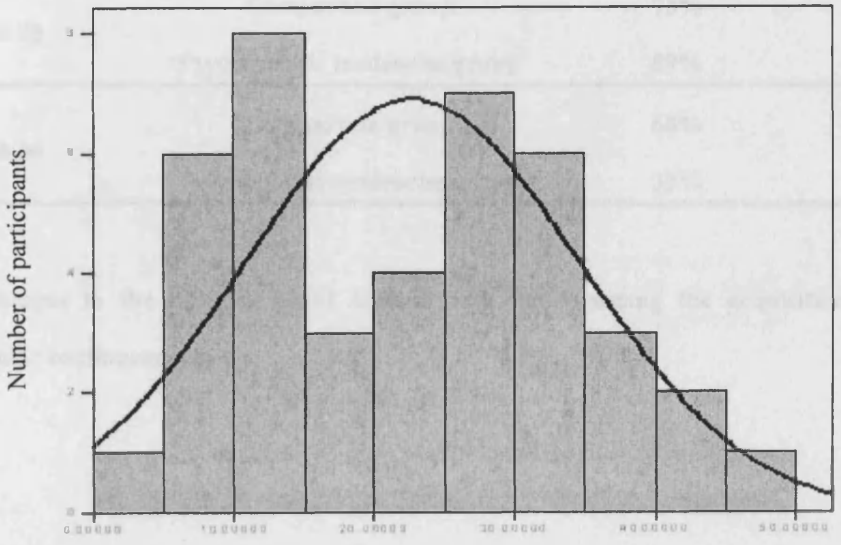
table 4.1: Mean age, BPVS score, and APSD and ADHD ratings

Group	Age	BPVS	APSD	ADHD
Psychopathic tendencies group (n=18)	12.53	90.06	29.30*	31.64*
	(2.01)	(11.05)	(2.80)	(8.60)
Comparison group (n=22)	13.03	88.09	11.77	15.35
	(1.87)	(8.61)	(2.89)	(9.18)

* $P < 0.001$ (standard deviations in parentheses)

Key to table 4.1.: APSD = Antisocial Process Screening Device (maximum score = 40); BPVS = British Picture Vocabulary scale; ADHD = attention-deficit hyperactivity disorder (maximum score = 54); n = number of participants.

figure 4.4: A distribution graph of DuPaul ADHD scores



4.4.1: Pass rate data

See table *4.2.* for the percentage pass rates for the children with psychopathic tendencies and the comparison children for the acquisition and reversal phases for each contingency. Chi-square tests revealed significant differences between groups only for acquisition in the 70-30 condition; χ^2 (1, N = 40) = 4.82, $P < 0.05$) (Chi-square/Fisher's exact test found no differences between groups for any other contingency at either phase).

table 4.2: Percentage pass rates in acquisition and reversal phases for each contingency

Contingency	Group	Phase	
		Acquisition	Reversal
100-0	Comparison group	86%	89%
	Psychopathic tendencies group	100%	89%
90-10	Comparison group	86%	79%
	Psychopathic tendencies group	74%	71%
80-20	Comparison group	73%	75%
	Psychopathic tendencies group	89%	47%
70-30	Comparison group	68%	80%
	Psychopathic tendencies group	33%	50%

(percentages in the reversal phase include only those passing the acquisition for that particular contingency)

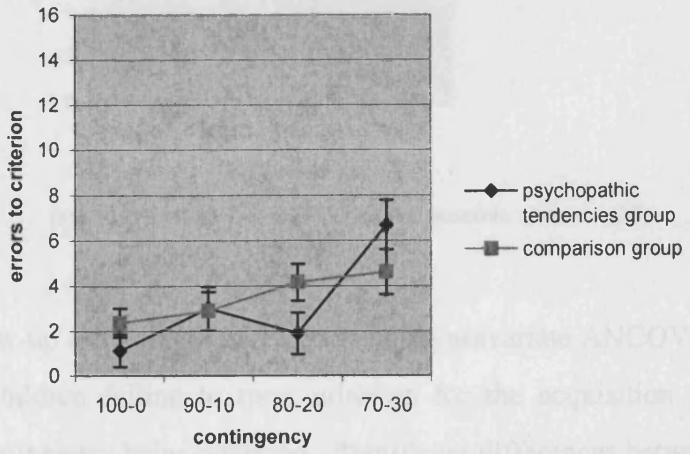
4.4.2: Error data

Errors to criterion (total errors made prior to reaching the criterion of 6 consecutive correct responses) were analysed using 2 repeated measures ANCOVAs. Separate analyses were conducted for phase 1 (Acquisition of the discrimination) and phase 2 (Reversal of the discrimination) as fewer participants were included in the reversal phase analyses. Participants were grouped according to scores on the APSD (psychopathic tendencies/comparisons). ADHD group membership (high/low) was the covariate. Probabilistic contingency (100-0, 90-0, 80-20, 70-30) was the within subjects factor.

Acquisition of the discrimination

As expected, neither the main effect of group ($F(1,37) < 1$; *NS*) nor the interaction between group and contingency ($F(3, 111) = 2.31$; *NS*) were significant (see figure *4.5.*). There was, however, a significant main effect ($F(3,111) = 2.78$; $P < 0.05$) and linear contrast ($F(1, 37) = 8.90$; $P < 0.005$) for contingency. ADHD was not a significant covariate ($F(1,37) < 1$; *NS*).

figure 4.5: Errors to criterion made by children in the acquisition phase

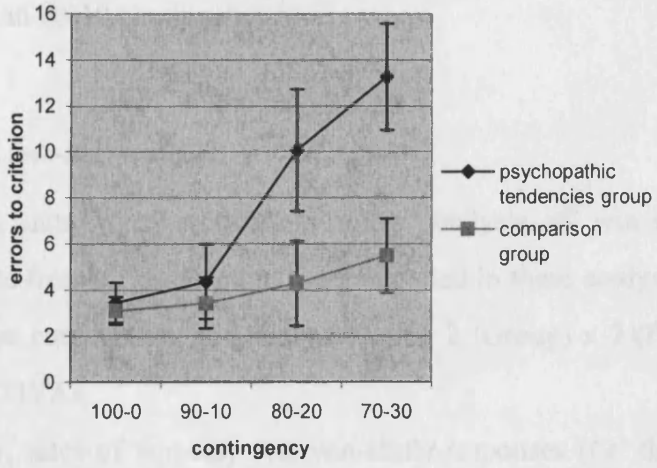


(standard error bars) (Maximum possible errors = 20)

Reversal of the discrimination

Those children not passing *all* four of the contingencies were excluded from the repeated measures ANCOVA (n = 25) assessing errors made in the reversal phase. There was a significant effect of group ($F(1,12) = 4.77$, $P < 0.05$) indicating that psychopathic tendencies group made more errors overall than comparison group. There was a also significant main effect of contingency ($F(3,36) = 5.63$, $P < 0.005$) and a significant linear contrast for contingency ($F(1,12) = 13.61$, $P < 0.005$). Further, the group by contingency interaction was significant ($F(3,36) = 2.98$, $P < 0.05$), as was the linear contrast ($F(1,12) = 6.99$, $P < 0.05$). ADHD was not a significant covariate ($F(1,12) < 1$; *NS*). In short, the psychopathic tendencies group made more errors than comparison group and the rate of errors increased as the reward: punishment ratio reduced across contingencies (see figure *4.6.*).

figure 4.6: Errors to criterion made by children in the reversal phase



(standard error bars) (Maximum possible errors = 20)

Follow-up analyses involved performing univariate ANCOVAs, excluding only those children failing to meet criterion for the acquisition phase for the particular contingency being analysed. Significant differences between error rates of the two groups were revealed in the 90-10, 80-20 and 70-30 contingencies; 100-0 ($F(1,34) < 1$, *NS*), 90-10 ($F(1,29) = 2.33$, $P < 0.05$, 1-tailed), 80-20 ($F(1,30)$

= 4.53, $P < 0.05$) and 70-30 (F(1,18) = 4.60, $P < 0.05$). ADHD was not a significant covariate in any of the analyses.

Consecutive perseverative errors (from the point at which the reward contingencies reversed) were also analysed using a 2 (group) x 4 (contingency) repeated measures ANCOVA (including only those children passing *all* four of the contingencies). Neither the main effect of group (F(1,12) = 2.11; *NS*), contingency (F(3,36) = 1.23; *NS*) nor group by contingency interaction (F(3,36) = 2.03; *NS*) were significant; both groups made equal rates of perseverative errors across all four contingencies.

Non-reversing pairs

Performance on the non-reversing pairs, with contingencies 90-10 and 70-30, was assessed using a 2 group by 2 contingency ANCOVA. The main effect of group was not significant (F(1,37) < 1, *NS*). There was a significant main effect for contingency (F(1,37) = 6.13, $P < 0.05$); both groups made more errors in the 70-30 rather than 90-10 contingency.

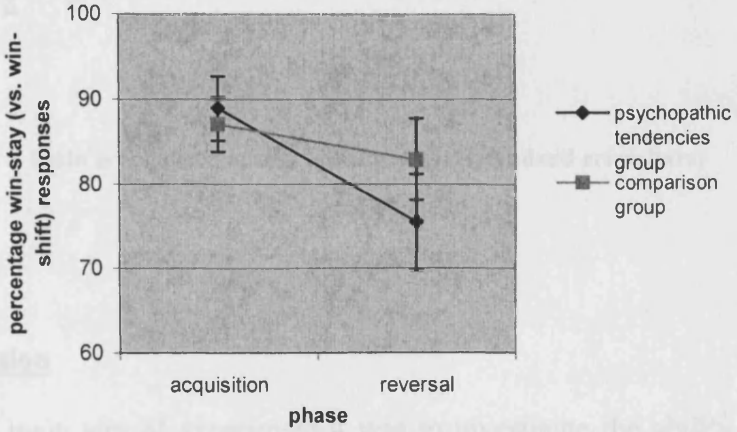
4.4.3: Win-stay, lose-shift analysis

Finally, data were recoded to allow analysis of win-stay, lose-shift strategies. Data from all participants were included in these analyses. Data were collapsed across contingency and analysed using 2 (Group) x 2 (Phase) repeated measures ANCOVAs.

Initially, rates of win-stay (vs. win-shift) responses (*i.e.* the participant's behaviour immediately following a *correct*, rewarded response) were assessed. As different participants made different numbers of correct responses, a win-stay percentage was calculated; [N. of win-stay/(N. of wins-stay + N. of win-shift) x 100]. Whilst the main effect of group was not significant (F(1,36) < 1, *NS*), there was a significant effect of phase (F(1,36) = 11.48, $P < 0.005$). The interaction between group and phase was also significant (F(1,36) = 2.80, $P < 0.06$, 1-tailed) (see figure *4.7.*). ADHD was not a significant covariate (F(1,36) < 1, *NS*).

Follow-up within-group ANOVAs revealed that the proportion of win-stay vs. win-shift responses decreased significantly from the acquisition phase to the reversal phase in the children with psychopathic tendencies ($F(1,16) = 10.71$, $P < 0.005$), but not in the comparison children ($F(1,21) = 3.92$, *NS*).

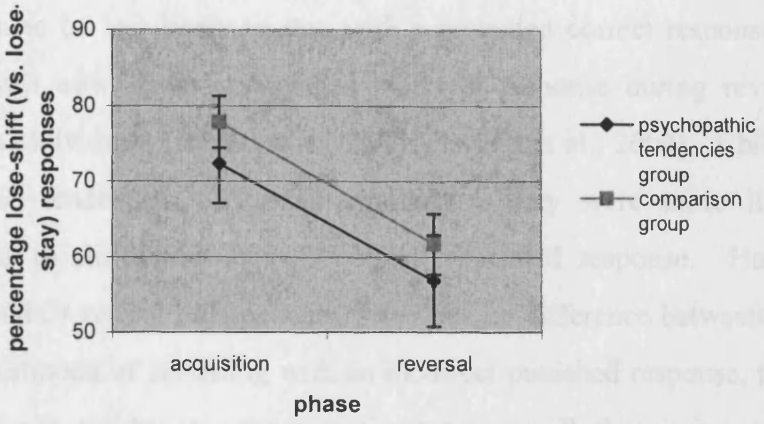
figure 4.7: Percentage win-stay (vs. win-shift) responses in the acquisition and reversal phases



(data is collapsed across contingencies) (standard error bars)

Following this, rates (expressed as percentages) of lose-shift (vs. lose-stay) responses (*i.e.* the participant's behaviour immediately following an *incorrect*, punished response) were examined; [N. of lose-shift/(N. of lose-shift + N. of lose-stay) x 100]. Neither the main affect of group, nor the group by phase interaction were significant, both ($F(1,15) < 1$, *NS*) (see figure *4.8.*). The main effect of phase was significant ($F(1,15) = 10.38$, $P < 0.01$).

figure 4.8: Percentage lose-shift (vs. lose-stay) responses in the acquisition and reversal phases



(data is collapsed across contingencies) (standard error bars)

4.5: Discussion

The main aim of experiment 4 was to investigate the ability of children with psychopathic tendencies to perform reversal learning under differing contingency conditions. Specifically, experiment 4 aimed to assess the effects of salience of the contingency change on performance, and to investigate whether any group effects were significantly related to psychopathic tendencies after the variation due to level of ADHD had been removed. In line with predictions, the children with psychopathic tendencies showed increasing reversal learning impairment, relative to the comparison group, as the salience of the contingency change decreased. This deficit was observed even when the impact of level of ADHD was statistically controlled.

Previous studies of extinction and reversal learning in children with psychopathic tendencies have found impairment when using both the Iowa gambling task and Newman's card extinction task but not when assessing the reversal learning phase of the ID/ED shift paradigm (Blair et al., 2001a; Fisher and Blair, 1998; O'Brien and Frick, 1996).

An alternative method of analysing the data was to compare the likelihood that participants would shift away from a correct, rewarded response, or stay with an incorrect, punished response. Patients with lesions to OFC/ ventral PFC have been shown to be less likely to stay with a rewarded correct response and less likely to shift away from a punished incorrect response during reversal than comparison individuals (Berlin et al., 2004; Hornak et al., 2004). Children with psychopathic tendencies performed similarly - they were more likely than comparisons to shift away from a correct, rewarded response. However, in contrast to OFC/ ventral PFC patients, there was no difference between groups in terms of likelihood of persisting with an incorrect punished response, that is, the impairment was not due to a perseverative tendency. Rather, it appears that the children with psychopathic tendencies were impaired in their ability to maintain new stimulus-response associations following reversal away from the newly incorrect stimulus.

Finally, in contrast to Itami and Uno (Itami and Uno, 2002), no significant impact of the ADHD covariate was observed. Neuro-anatomically ADHD has been linked to dysfunction in the right frontal cortex, anterior cingulate and basal ganglia (*e.g.* Casey et al., 1997; Castellanos et al., 1996; Giedd et al., 2001; Swanson et al., 1998), whilst psychopathy has been linked with dysfunction in amygdala and orbital/ ventrolateral frontal cortex (*e.g.* Blair et al., 2001a; Blair, 2003a; Blair et al., 2004; Blair et al., 1999; Damasio, 1994; LaPierre et al., 1995; Mitchell et al., 2002; Patrick, 1994). It is possible that Itami and Uno (Itami and Uno, 2002) found group differences because of their participant selection criteria. They did not screen for psychopathic tendencies, and it is possible, given high co-morbidity rates between psychopathic tendencies and ADHD (Colledge and Blair, 2001), that at least some children in their ADHD group had psychopathic tendencies. Alternatively they may have recruited an extreme ADHD group, with more severe symptomatology than observed in this experiment. Given the high co-morbidity of ADHD with CD/psychopathic tendencies (Babinski et al., 1999; Barry et al., 2000; Biederman et al., 1991; Colledge and Blair, 2001; Hinshaw, 1987; Lynam, 1996; Taylor et al., 1986) and the interest in determining which

deficits are specific to which disorder, this issue should be re-investigated in further work.

4.5.1: Summary and conclusions

In summary, this experiment investigated the performance of children with psychopathic tendencies on a novel probabilistic reversal learning task. In line with predictions, children with psychopathic tendencies showed dysfunctional reversal learning only in the probabilistic conditions. Also the performance decrement shown by children with psychopathic tendencies was an inverse function of the salience of the contingency change, that is, as contingencies became more probabilistic the children with psychopathic tendencies committed increasing rates of errors. Finally, it appeared that the reason for the reversal learning impairment was due to an inability to maintain new responses following contingency reversal.

4.6: Experiment 5

As discussed above, previously, inconsistencies have been observed in the literature detailing performance of children with psychopathic tendencies and adult psychopaths on reversal learning paradigms. It appeared that while adult psychopaths presented with difficulties in reversal learning regardless of salience of contingency change, children with psychopathic tendencies seemed able to successfully perform simple reversals, presenting with difficulties only on less salient reversals. In line with this observation, the results of experiment 4 provided further evidence that children with psychopathic tendencies are able to successfully perform simple reversals, and only present with impairment on less salient reversals. In addition, and somewhat consistent with the performance of patients with OFC lesions, the children with psychopathic tendencies were found to be more inclined to shift away from correct, rewarded responses than were comparisons. In contrast to children with psychopathic tendencies, adult psychopaths have reliably presented with impaired performance on reversal learning tasks (see section *4.2.*). They have shown difficulties on tasks assessing reversal and extinction regardless of the salience of contingency change. In short, it appears that the reversal learning impairment may be more pronounced in adult psychopaths than it is in children with psychopathic tendencies.

4.6.1: Summary of Aims

Experiment 5 attempted to assess the performance of a group of adult individuals with psychopathy on a novel probabilistic reversal learning paradigm. Specifically experiment 5 aimed to replicate previous literature with adult psychopaths and also to investigate their behavioural strategy (*i.e.*, win-stay and lose-shift).

4.7: Methods

4.7.1: Design

The independent variables were: group (psychopaths/comparisons); the two phases (acquisition of the discrimination/reversal of the discrimination); and the two reinforcement contingencies (100-0/80-20). The dependent variable was errors to criterion (see section *4.7.4.* for details).

4.7.2: Participants

The sample was made up of 38 men from a pool of 200 men incarcerated in a forensic institution in the London area. 30 of the participants were Caucasian, 1 Asian, and 8 Afro-Caribbean (1 Asian and 2 Afro-Caribbean participants were in the comparison group, and 6 Afro-Caribbean participants were in the psychopathic group). Files were pre-screened to exclude individuals whose psychiatric reports revealed a diagnosis for psychosis, organic brain damage, or neurological disorder.

In accordance with the literature and the guidelines of the PCL-R (Hare, 1991; Hare, 2003), individuals with a score of 30 or above on the PCL-R were assigned to the psychopathy group (n = 20), whilst those with a score of 20 or less were assigned to the comparison group (n = 18). Written consent was obtained from each participant who participated in the experiment, and all were informed that they were free to withdraw from the experiment at any time.

4.7.3: Measures

Raven's Advanced Progressive Matrices, Set I (Raven, 1976).

Set 1 of the Raven's Advanced Progressive Matrices was used to provide an estimation of full-scale intelligence quotient (IQ). This measure is not dependent on the participant's ability to read.

Psychopathy Checklist-Revised (PCL-R; Hare, 1991; Hare, 2003).

Participant's scores for each item were the averages assigned by two independent raters. Pearson's correlations of the two were; for total PCL-R score $r2 = 0.98$ ($P < 0.001$). The agreement between the two raters for diagnostic group (psychopathic vs. comparison) was 100%.

4.7.4: Probabilistic Reversal learning Task.

The task was programmed in Visual Basic (6.0) and was presented on a Dell Laptop computer. Stimuli comprised 12 line drawings of animals (Snodgrass and Vanderwart, 1980) which had each been shaded in a different colour (see figure 4.2). Stimuli measured 4cm by 4cm and were presented on a grey background.

Stimuli were assigned into pairs randomly at the beginning of the task. On each trial, a pair of stimuli was presented. Stimulus locations were assigned randomly on each trial (there were 16 possible locations). The participant had to choose one of the stimuli by clicking on it with the mouse. Upon choosing they would receive either positive ('you win 100 points') or negative ('you lose 100 points') feedback (for 1000 ms) according to the reinforcement contingency of that pair. One of the animals in each pair was always more likely than the other to be rewarded rather than punished. Participants began the task with 0 points. A running total of points was visible at the bottom of the screen only during the feedback display. Trials were self-paced.

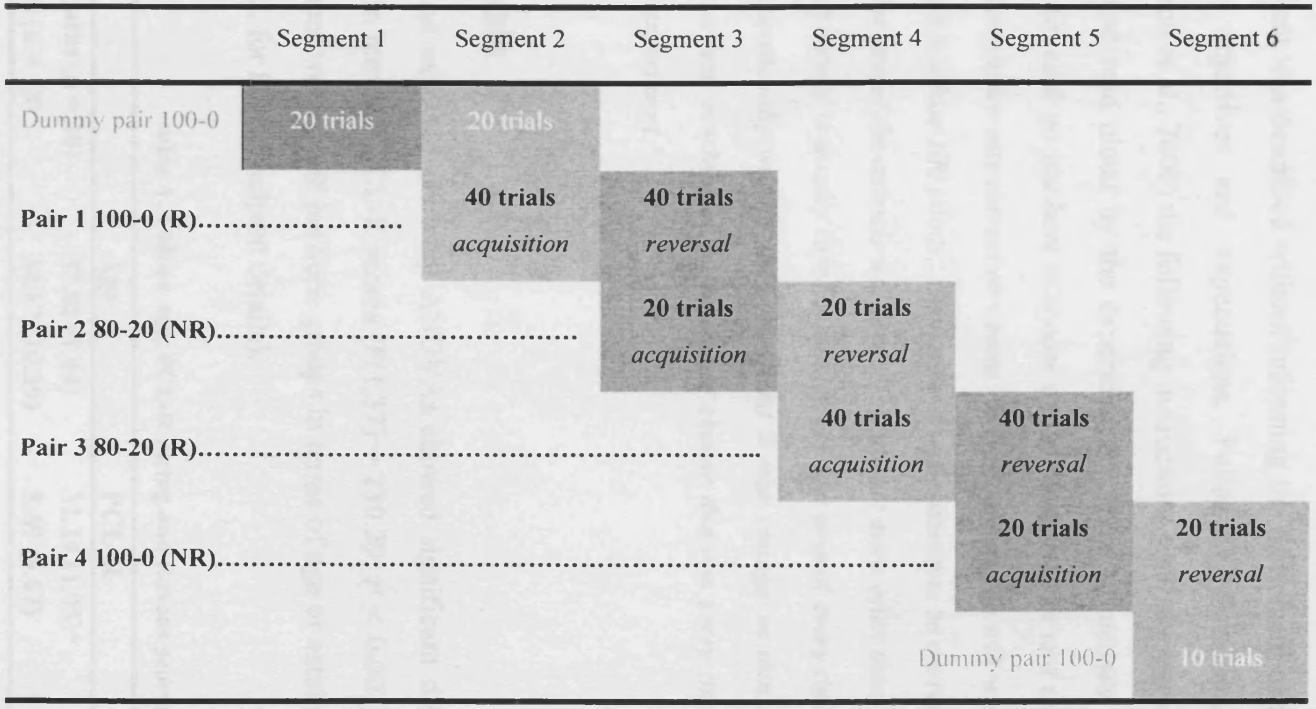
There were 2 reinforcement contingencies; 100-0 (simple) and 80-20 (probabilistic). In the 100-0 condition the 'correct' stimulus was always rewarded and the 'incorrect' stimulus was always punished. In the 80-20 condition the 'correct' stimulus was rewarded on 8 out of every 10 occasions and punished on 2 out of every 10 occasions. Inversely, the 'incorrect' stimulus was be punished on 8 out of every 10 trials and rewarded on 2 out of every 10 trials. The order of probabilistic feedback was randomised within the program.

There were four different experimental pairs of stimuli: two of which changed contingency (reversing pairs; R) and two of which did not (non-reversing pairs; NR) (see figure 4.9 for a detailed task diagram). The 2 reversing pairs had contingencies 100-0 and 80-20 and the two non-reversing pairs had contingencies 100-0 and 80-20. The reinforcement contingency of the reversing pairs remained constant for 40 trials (phase 1; acquisition of the discrimination). Upon completing 40 trials the reinforcement contingency the reversing pairs was reversed (phase 2; reversal of the discrimination). Thus the previously correct stimulus became the incorrect stimulus and the previously incorrect stimulus now became the correct stimulus. The contingencies of the non-reversing pairs remained the same for the entire 40 trials that they were presented for.

Rather than learn about pairs of stimuli serially (*e.g.*, all 40 trials of 100-0 contingency followed by all 40 trials of 90-10 contingency, and so on), participants had concurrent experience with two different pairs of stimuli at all times (*i.e.* trials from two pairs were presented alternately). This was done in order to increase task difficulty; serial presentation might have allowed the participant to more easily calculate that after a set amount of trials, many of the pairs changed contingency. The presentation order of the pairs was staggered so that the reversals did not coincide. Initially a practice pair (non-reversing with a 100-0 reinforcement contingency) was presented alone for 20 trials (to introduce participants to probabilistic reinforcement) and then alternately with the first experimental pair. Towards the end of the task a 'dummy' pair (non-reversing with a 100-0 reinforcement contingency) was introduced and presented on alternate trials with the final experimental pair. Thus each experimental pair was presented on alternate trials with another pair of stimuli. All participants received 290 trials regardless of performance.

For the purposes of analysis a learning criterion of 8 consecutive correct responses was imposed in both phases. Thus participants had to choose the correct stimulus in each pair eight times consecutively before they had successfully passed that phase of the task. If participants did not meet the learning criterion total errors made were analysed.

figure 4.9: A diagrammatic representation of the probabilistic reversal learning task



The orders presentation of pairs 1-4 was randomized. In each segment trials from pairs were presented alternately

Key to figure 4.9.: R = reversing, NR = non-reversing

4.7.5: Procedure

Each participant was tested individually in a quiet room allocated by the institution. Subsequent to the administration of the Raven's Progressive Matrices by the experimenter, the participants completed the reversal learning task. The experiment was described without informing the participant of the investigation's specific objectives and expectations. Following Swainson and colleagues (Swainson et al., 2000) the following instructions were presented on the computer screen and read aloud by the experimenter: '*Pairs of animals will appear on the screen. On each go you have to choose one of these animals and the computer will tell you if your choice was correct or wrong. If it is correct you will win 100 points. If it is wrong you will lose 100 points. Each animal will sometimes be correct and sometimes be wrong, but one of the animals will tend to be correct more often than the other one. Find out which animal is usually correct, and choose that animal every time. Stick with it even if it is occasionally wrong. At some point it may change so that the other animal is usually correct, in which case you should choose that one every time. Press 'begin' to start the experiment.*'

4.8: Results

As expected one-way ANOVAs showed significant differences between groups in terms of PCL-R scores ($F(1,37) = 230.29$, $P < 0.001$)). No significant differences were found between groups in terms of age or estimated IQ score (see table *4.3.* for full participant details).

table 4.3: Mean age, PCL-R rating and Ravens scores

Group	Age	PCL-R	Ravens
Psychopaths (n = 20)	**37.80 (7.64)**	**32.24 (1.95)***	**8.05 (1.99)**
Controls (n = 18)	**34.17 (10.39)**	**8.59 (6.67)**	**8.22 (1.86)**

*** $P < 0.001$**
(standard deviations in parentheses)

Key to table 4.3.: PCL-R = Psychopathy Checklist – Revised (maximum score = 40); Ravens = Raven's Progressive Matrices (maximum score = 12); n = number of participants.

4.8.1: Failure rate data

See table *4.4.* for the percentage pass rates for the adult psychopaths and the comparison adults for the acquisition and reversal phases for both contingencies. Chi-square/Fisher's Exact tests revealed no significant differences between groups in terms of pass rates.

table 4.4: Percentage pass rates for adult psychopaths and comparison adults in the acquisition and reversal phases for each contingency

Contingency	Group	Phase	
		Acquisition	Reversal
100-0	Comparison group	100%	100%
	Psychopathic group	100%	95%
80-20	Comparison group	94%	94%
	Psychopathic group	85%	82%

(percentages in the reversal phase include only those passing the acquisition for that particular contingency)

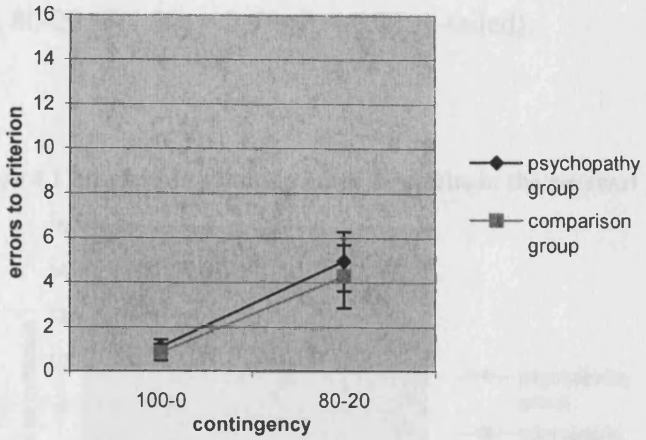
4.8.2: Error data

Errors to criterion (total errors made prior to reaching the criterion of 8 consecutive correct responses) were analysed using 2 repeated measures ANOVAs. Separate analyses were conducted for phase 1 (Acquisition) and phase 2 (Reversal) errors. Participants were grouped according to scores on the PCL-R (psychopathic/ comparisons). Probabilistic contingency (100-0 and 80-20) was the within subjects factor.

Phase 1: Acquisition of the discrimination

As expected neither the main effect of group nor the interaction between group and contingency were significant: both ($F(1,36) < 1$; *NS*). There was a significant main effect of reinforcement contingency ($F(1, 36) = 18.57$; $P < 0.001$). Thus, as predicted, all participants made more errors learning the 80:20 contingency than the 100:0 contingency (See figure *4.10.*).

figure 4.10: Errors to criterion made by adults in the acquisition phase



(standard error bars) (Maximum possible errors = 40)

Phase 2: Reversal Phase of the discrimination

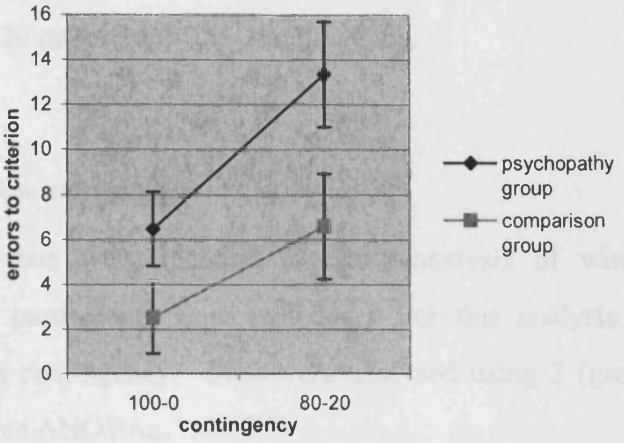
For the repeated measures ANOVA assessing errors made in the reversal phase, those individuals not meeting criteria for *both* of the contingencies were excluded from the analysis (n = 4). As predicted there was a significant main effect of group ($F(1,32) = 6.11$, $P < 0.05$); adults with psychopathy made significantly more reversal errors than control adults. There was also a significant main effect of contingency ($F(1,32) = 8.64$, $P < 0.01$), however the group by contingency interaction was not significant ($F(1,32) < 1$, *NS*) (see figure *4.11.*).

Follow-up univariate ANOVAs included all participants for the 100-0 contingency, and excluded four participants failing to meet criterion in the acquisition phase for the 80-20 contingency. These revealed significant group differences for both contingencies; 100-0 ($F(1,36) = 3.44$, $P < 0.05$, 1-tailed), and 80-20 ($F(1,32) = 4.19$, $P < 0.05$).

Consecutive perseverative errors were analysed from the point at which the reward contingencies reversed. There was a significant main effect for group ($F(1,32) = 4.01$; $P < 0.05$). Neither the main effect of contingency ($F(1,32) = 1.18$; *NS*) nor group by contingency interaction were significant ($F(1,32) < 1$; *NS*) (see figure *4.12.*). Follow-up univariate ANOVAs revealed that the group

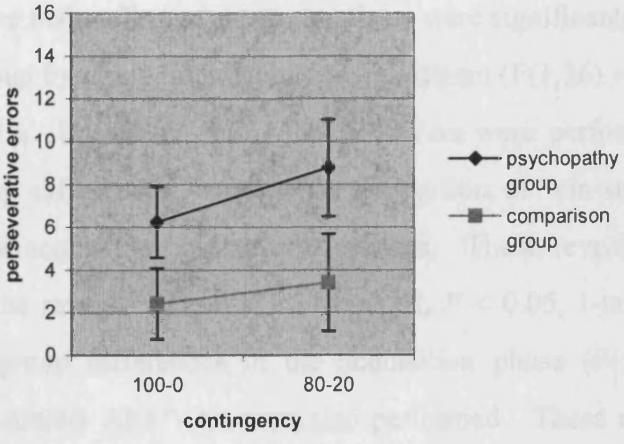
differences were significant for both contingencies: 100-0 ($F(1,36) = 2.56$, $P < 0.05$, 1-tailed); 80-20 ($F(1,32) = 2.79$, $P < 0.05$, 1-tailed).

figure 4.11: Errors to criterion made by adults in the reversal phase



(standard error bars). (Maximum possible errors = 40)

figure 4.12: Consecutive perseverative errors made by adults in the reversal phase



(standard error bars). (Maximum possible errors = 40)

Non-reversing pairs

Performance on the non-reversing pairs was also assessed. Neither the main effect of group ($F(1,36) = 3.10$, *NS*) nor the group by contingency interaction ($F(1,36) = 1.90$, *NS*) were significant. There was a significant main effect for contingency ($F(1,35) = 19.50$, $P < 0.001$); both groups made more errors in the 80-20 rather than 100-0 contingency.

4.8.3: Win-stay, lose-shift analysis

Finally, data were recoded to allow analysis of win-stay, lose-shift strategies. All participants were included. For this analysis responses were collapsed across contingency. Data were analysed using 2 (group) x 2 (phase) repeated measures ANOVAs.

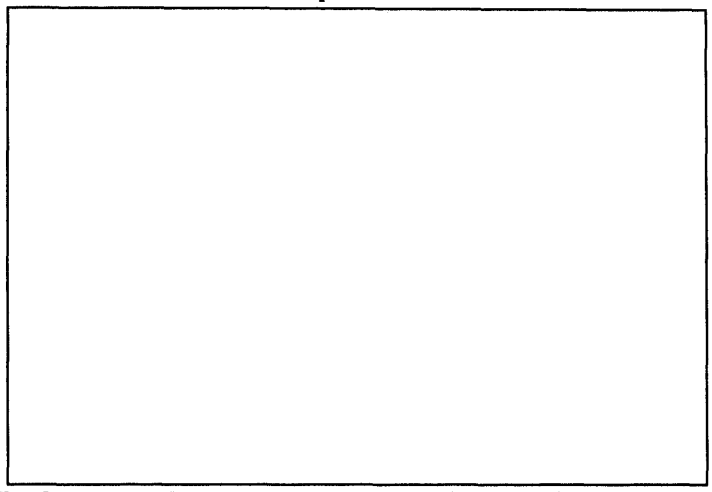
Initially, rates of win-stay (vs. win-shift) responses (*i.e.* the participant's behaviour immediately following a *correct*, rewarded response) were assessed. As participants made different numbers of correct responses, a win-stay percentage was calculated [N. of win-stay/(N. of wins-stay + N. of win-shift) x 100]. Neither the main effect of group nor phase were significant: both ($F(1,36) < 1$, *NS*). The group by phase interaction was significant ($F(1,36) = 6.16$, $P < 0.05$) (see figure *4.13.*). Univariate follow-up ANOVAs were performed in order to assess the group differences between the proportion of win-stay vs. win-shift responses in the acquisition and reversal phases. These revealed a significant effect only for the reversal phase ($F(1,34) = 3.02$, $P < 0.05$, 1-tailed); there were no significant group differences in the acquisition phase ($F(1,36) < 1$, *NS*). Further, within-subject ANOVAs were also performed. These revealed that the proportion of win-stay vs. win-shift responses decreased significantly from acquisition to reversal phases only for the individuals with psychopathy ($F(1,17) = 11.25$, $P < 0.005$); for the comparison group ($F(1,17) = 1.14$, *NS*).

figure 4.13: Percentage win-stay (vs. win-shift) responses in the acquisition and reversal phases

(the data are collapsed across contingencies) (standard error bars)

Following this, rates (expressed as percentages) of lose-shift (vs. lose-stay) responses (*i.e.* the participant's behaviour immediately following an *incorrect*, punished response) were examined [N. of lose-shift/(N. of lose-shift + N. of lose-stay) x 100]. Neither the main effect of group nor the group by phase interaction were significant: both ($F(1,36) < 1$, *NS*) (see figure *4.14.*). The main effect of phase was significant ($F(1,36) = 46.86$, $P < 0.001$). However, this was principally because punishment during the reversal phases of the paradigm had no predictive power with respect to the participant's subsequent response; participants were at chance as to whether they would make the same incorrect response or the alternative correct response following a punished incorrect response.

figure 4.14: Percentage lose-shift (vs. lose-stay) responses in the acquisition and reversal phases



(The data are collapsed across contingencies) (standard error bars)

4.9: Discussion

Experiment 5 examined the ability of adult individuals with psychopathy to perform a novel probabilistic reversal learning task. The task involved one simple contingency reversal and one probabilistic contingency reversal. As expected the individuals with psychopathy performed similarly to comparisons in the acquisition phase of the task, whilst, in the reversal phase, the individuals with psychopathy committed a greater number of errors in both contingencies. The salience of contingency change had a comparable impact on both groups across phases.

Results were in accordance with previous observations in the literature, indicating that adult individuals with psychopathy present with impairment in reversal learning regardless of the salience of contingency change (Mitchell et al., 2002). Thus group error rates were significantly different for both the highly salient (100-0) and less salient (80-20) contingency reversals. One interesting feature of the current results (and in contrast to those obtained in experiment 4) is that there was no significantly greater impact of reduced contingency salience on

reversal learning in the individuals with psychopathy than the comparison individuals.

On the basis of previous human neuropsychological work (Berlin et al., 2004; Hornak et al., 2004), it was predicted that individuals with psychopathy would be less likely to stay with a rewarded correct response and less likely to shift away from a punished incorrect response during reversal than comparison individuals. The first of these predictions was confirmed. The second was not; surprisingly, both groups were as likely to shift away from, as to stay with, a punished incorrect response during reversal. These results are in accordance with those obtained in experiment 4 with children with psychopathic tendencies.

4.9.1: Summary and conclusions

In summary, this experiment investigated the performance of adult individuals with psychopathy on a probabilistic reversal learning task. In line with predictions, adults with psychopathy performed the acquisition phases successfully. Also as predicted, they showed a performance deficit in both simple and probabilistic conditions relative to comparison individuals. Finally, as in experiment 4, it appeared that an inability to maintain new responses following contingency reversal contributed to the reversal learning impairment.

4.10: General Discussion

Chapter 4 presented two experiments. Experiment 4 assessed the performance of children with psychopathic tendencies and comparisons on a novel reversal learning task. As predicted, children with psychopathic tendencies presented with increasing impairment as the salience of contingency change decreased, but only in the probabilistic conditions. Experiment 5 assessed performance of adults with psychopathy and comparison individuals on a modified version of the reversal learning task. The adults with psychopathy presented with impairment relative to comparisons in both the simple and probabilistic conditions. In addition children with psychopathic tendencies and adult psychopaths were found to be more likely to move away from a correct, rewarded response in the reversal phase of the tasks, relative to comparisons. These findings strengthen the claim that there exists some impairment within OFC/ ventrolateral PFC in children with psychopathic tendencies and adults with psychopathy. In addition experiments 4 and 5 provide further evidence in line with the suggestion that the impairment is greater in adult psychopaths than in children with psychopathic tendencies.

4.10.1: Dissociation between acquisition and reversal phases

Experiments 4 and 5 demonstrated that children with psychopathic tendencies and adult psychopaths show impairment, relative to comparisons, only in the reversal phase of reversal learning tasks; importantly the initial acquisition of the discrimination remains intact. Thus, whilst the ability of these individuals to perform reversal learning is compromised, they are able to perform object discrimination learning. This dissociation is of importance particularly as the current results cannot be attributed to a task difficulty effect. For example, the adult psychopaths found the reversal of the 100-0 contingency to be more difficult than acquisition of the 80-20 contingency, whilst the converse was true for comparison individuals (as evident by error rates, see figures *4.10.* and *4.11.*). In other words, the individuals with psychopathy showed impairment relative to the

comparison individuals on the condition that was easier for the comparison individuals. This dissociation also precludes a motivational account of task performance. In particular, a motivation-based account would need to explain how reduced enthusiasm would give rise to impairment on the 'easy' 100-0 reversal and the 'difficult' 80-20 reversal but not the 'medium difficulty' 80-20 acquisition.

4.10.2: Implications of these Results for the Characterization of Psychopathy

Impairments in reversal learning have been observed in other psychiatric disorders that have been associated with a heightened risk of frustration or threat-based reactive aggression, such as intermittent explosive disorder (Best et al., 2002) and paediatric bipolar disorder (Gorrindo et al., in press). Moreover, as noted in section *1.3.3.*, patients with lesions of the OFC/ ventrolateral PFC often present with increased rates of reactive aggression and also impaired ability to perform reversal learning. Indeed, it has been suggested that the presence of irritability and reactive aggression in certain psychiatric disorders might be a marker for dysfunction within OFC/ ventrolateral PFC (Blair, 2001; Blair, 2004; Leibenluft et al., 2003). The current data provide further support for the contention that there exists such dysfunction in individuals with psychopathy (Blair, 2003a; Blair, 2004; Blair et al., 2001a; Mitchell et al., 2002). Interestingly, and similar to individuals with OFC/ventrolateral damage, both children with psychopathic tendencies and adult psychopaths were observed to return to the previously correct stimulus *after* successful behavioural reversal has been achieved (Berlin et al., 2004; Hornak et al., 2004). Specifically, this manifested as an increased tendency to shift stimulus choice directly following a rewarded, correct response. However, in contrast to individuals with OFC/ventrolateral PFC damage, both children with psychopathic tendencies and adult psychopaths were as likely as controls to shift away from, as to stay with, a punished incorrect response during reversal. In contrast OFC/ ventrolateral PFC patients present with a marked tendency to perseverate towards a previously rewarded (and now punished) stimulus (Berlin et al., 2004; Fellows and Farah, 2003; Hornak et al.,

2004; Rolls et al., 1994). It thus appears that the reversal learning deficit observed in children with psychopathic tendencies and adult psychopaths must be characterised at least somewhat differently to patients with OFC/ ventrolateral PFC lesions. Although adult psychopaths do perseverate somewhat, it appears that the deficit in children with psychopathic tendencies and adult psychopaths may be at least partly due to a tendency to return to a previously rewarded stimulus *after* they have successfully reversed responses.

Following Blair (2004), in order to successfully achieve reversal learning or extinction, the participant must encode the motivational significance of cues and detect whether the expected outcomes differ from the actual outcomes. On occasions when the incentive value of the expected outcomes differs from the actual outcomes the individual must alter their behaviour. Thus reversal learning must require: (1) the representation of the expected reward/punishment; (2) the representation of the actual reward/punishment; (3) the ability to detect contingency violations; and (4) the ability to alter stimulus-response associations following the detection of these violations. Data from chapter 4 indicates that children with psychopathic tendencies and adult psychopaths may present with difficulty in stage 4. Further, on the basis of these data, it appears that stage 4 may be decomposed further into separable components; the ability to alter and the ability to *maintain* new stimulus-response associations.

The results of experiments 4 and 5 were broadly in line with each other. However, one difference was that the children with psychopathic tendencies were unimpaired on the simple (100-0) reversal whereas the adults with psychopathy were. That is, it appears that the reversal learning impairment in adult psychopaths is greater than that observed in children with psychopathic tendencies. It may be suggested that this age-group difference could be due to external factors, for example the different instruments used to identify psychopathy in the two age groups. However it must be noted that children with psychopathic tendencies and adult psychopaths, identified using the same measures used in experiments 4 and 5, have performed comparably in other investigations (e.g. in tasks aiming to assess the integrity of the amygdala;

Aniskiewicz, 1979; Blair, 1999; Blair and Coles, 2000; Blair et al., 2001c; Blair et al., 1997; House and Milligan, 1976; Stevens et al., 2001; Sutker, 1970). Thus is seems unlikely that the adults were simply a more extreme group.

It remains unclear why the reversal learning impairment would be greater in adults with the disorder. However, several possibilities may be considered. Firstly, it could be developmentally independent of the core pathology, that is, (following the IES model; Blair, 2004), amygdala pathology. For example, the genetic anomalies associated with psychopathy (Viding et al., 2005) might affect the development of the amygdala and OFC/ ventrolateral PFC independently of one another. Secondly, it could be dependent upon dysfunction within other areas. A primary candidate structure allowing for such an impact would be the amygdala. It is known that the amygdala and OFC are interconnected (Amaral et al., 1992; Rolls, 1997). A reduction in afferent input from the amygdala could potentially have negative consequences for the responsiveness of the OFC, leading to reduced sensitivity to contingency change in individuals with psychopathy as they age. Indeed, rodent data indicates that damage within the basolateral nucleus of the amygdala leads to reduced firing within OFC (O'Doherty, 2003; Schoenbaum et al., 2003). Thirdly, the greater OFC/ ventrolateral PFC dysfunction seen in the adults may be a secondary consequence of some of the behavioural characteristics of psychopathy. For example, one of the criteria of psychopathy, stimulation seeking, is often associated with drug use (Hare, 1991). Indeed, studies have indicated that psychopathy is associated with higher rates of drug abuse, and poly drug use (*e.g.* Hemphill et al., 1994; Smith and Newman, 1990). Bechara and colleagues observed that the majority of substance dependent patients performed the Iowa gambling task within the range of patients with ventromedial lesions (Bechara et al., 2001). Further, using a novel decision-making task, Rogers et al (Rogers et al., 1999b) assessed the quality of decision-making and deliberation time of individuals with focal OFC damage, and individuals who abused amphetamine or opiates. All three groups showed impaired performance on the task relative to comparison groups. Furthermore, chronic amphetamine abusers showed a pattern of sub-optimal

decision-making, that was similar to the pattern shown by the OFC patients, and which correlated with their years of abuse (Rogers et al., 1999b). Given the neuro-cognitive impairments associated with chronic drug abuse, and the data suggesting higher rates of abuse and dependence among psychopathic individuals, it is possible that the greater reversal learning deficit seen in adults with psychopathy, relative to children with psychopathic tendencies, is a secondary consequence of the stimulus seeking behaviour characteristic of the disorder.

4.10.3: Implications of these Results for the Theories of Psychopathy

Of the six theories discussed in the previous chapter, the current results are relevant for the fear dysfunction hypotheses (Cleckley, 1976; Eysenck, 1964; Fowles, 1988; Gray, 1987; Lykken, 1995; Mealey, 1995; Patrick, 1994; Pichot, 1978; Trasler, 1978; Trasler, 1973), the response-set modulation (RM) hypothesis (Newman, 1998; Patterson and Newman, 1993) and the integrated emotion systems (IES) hypothesis (Blair, 2004).

The dissociation between task performance in the acquisition and reversal phases (discussed in section *4.10.1.*) is potentially problematic for two of the main positions on psychopathy. The fear and the RM positions stress a general impairment in processing punishment-related information and an inability to shift away from the goal of responding to gain reward to the peripheral punishment information as the primary impairments in psychopathy respectively. According to these formulations, both positions should predict impairment during acquisition *and* reversal. The fear dysfunction position asserts that individuals with psychopathy are impaired in processing punishment-related information (Lykken, 1995; Patrick, 1994). This suggests that impairment in both phases should be predicted, as both require appropriate learning on the basis of punishment-related information. In the least, the fear positions require adaptation, to specify that the processing of punishment information is only impaired in the context of tasks reliant on the amygdala. According to the RM hypothesis, the poor performance of individuals with psychopathy on emotional learning tasks such as passive avoidance learning and the one-pack card playing tasks is related to their inability

to shift attention from their goal of responding to gain reward, to the peripheral punishment information. Again this position should predict impairment in both the acquisition and reversal phases of reversal learning tasks as, for successful performance both require the participant to shift attention from the goal of gaining reward to process the 'peripheral' punishment information. The results of experiments 4 and 5 therefore also suggest that the RM hypothesis needs adaptation. In the very least, the current results suggest that psychopaths are able to shift their attention away from the goal of responding to gain reward to peripheral punishment information in the context of an object discrimination paradigm (*i.e.* during the acquisition phase). However, the RM hypothesis has yet to provide a reason why the putative attentional difficulties might be confined to reversal of an object discrimination (or acquisition of a stimulus-reinforcement association, as indexed by passive avoidance learning).

The results of experiments 4 and 5, however, are relatively successfully compatible with the IES model (Blair, 2004). According to this model, individuals with psychopathy present with dysfunction within amygdala and OFC/ ventrolateral PFC. Following Baxter and Murray (2002), object discrimination (*i.e.* stimulus-response learning) does not recruit the amygdala. However, OFC/ ventrolateral PFC are believed to play roles in the *modification* of stimulus-response associations (*e.g.* in reversal learning) (Blair, 2004). The IES model would therefore not expect that individuals with psychopathy would present with impairments in the acquisition (*i.e.* object discrimination) phase of the probabilistic reversal learning task. Previous data provided by the ID/ED task did not allow an adequate test of the contrasting predictions of the fear/RM and the IES. The object discrimination phases of the ID/ED task are considerably easier than the two reversal learning phases. Failure to find group differences might have simply reflected a floor effect: the task may have been too simple for both groups. The replications and extensions provided by experiments 4 and 5, however, lend further support to the IES model of psychopathy.

Within the model (see section *1.3.6.* and figure *1.1.*), there are two systems that are particularly relevant to reversal learning. The first consists of units that

code the expected level of reward following the commission of specific actions to specific stimuli. These units allow rapid decision making and are thought to be implemented by medial OFC. This second system is thought to be implemented by ventrolateral PFC. It codes if a contingency expectation is violated; *i.e.*, if the expected reward does not occur. Ventrolateral PFC is thought to be involved in preventing responses that are no longer appropriate. In a reversal learning task, the response activated as a result of the previously acquired stimulus-response association is no longer appropriate. In contrast to this formulation, however, on the basis of the results of experiment 3 it appears that medial OFC, rather than ventrolateral PFC is important in the detection of expectation violations (see section *3.5.*)

The IES model follows Dayan and colleagues in stressing the importance of temporal difference calculations (O'Doherty et al., 2004; Schultz et al., 1997; Sutton and Barto, 1981). The temporal difference error is the difference between the expected value associated with a stimulus/action and the actual value currently received with respect to that stimulus/action. In other words, unexpected rewards induce large positive temporal difference errors (initiating rapid learning). In contrast, absent highly expected rewards induce large negative temporal difference errors (initiating reversal learning/extinction). As regards the reversal learning task; for salient contingency changes, the previous action has always been rewarded and now is always punished. In this case the initial negative temporal difference errors following contingency change will be very large, (*i.e.* there will be a strong impetus for reversal learning/extinction). For less salient contingency changes, the expected reward value associated with the old action will be less (because it was inconsistently rewarded) and the current values are less likely to be punishments (because these are also inconsistent). In this case the initial negative temporal difference errors following contingency change will be smaller. It appears that children with psychopathic tendencies and adult psychopaths may be impaired in their response to these negative temporal difference errors.

The results of the win-stay, lose-shift analysis, specifically the increased tendency to for the children with psychopathic tendencies and adult psychopaths to win-shift, were rather surprising and are not currently successfully accounted for by either the fear, RM or IES models. It has been recently shown that patients with damage to orbital/ventrolateral PFC also present with an increased tendency to perform win-shift responses (Berlin et al., 2004; Hornak et al., 2004). However, whilst both children with psychopathic tendencies and adult psychopaths exhibited an increased tendency to shift stimulus choice directly following a rewarded, correct response, both groups were as likely to shift away from, as to stay with, a punished incorrect response during reversal. It thus appears that the reversal learning deficit observed in children with psychopathic tendencies and adult psychopaths must be characterised at least somewhat differently to patients sustaining lesions to OFC/ ventrolateral PFC.

The results of the win-stay, lose-shift analysis are difficult to account for by either the fear or RM positions. Both emphasize impaired processing following the receipt of punishment, however, on the basis of these results, it appears that the cause for reversal learning impairment in psychopathy is not simply due to insensitivity to punishment and thus a tendency to perseverate with a punished response. The IES model is also currently unable to account for these findings. In particular, the model states that the OFC codes expectations of reward and punishment, whilst ventrolateral PFC comprises units involved in detecting contingency violations and preventing motor responses that are no longer appropriate (Blair, 2004). Specifically, it is thought that, in a reversal learning task, the response activated as a result of the previously acquired stimulus-response association is no longer appropriate, and in healthy individuals the ventrolateral PFC acts to inhibit these responses. The results of the win-stay analysis suggests that this formulation requires modification, specifically, that it must also detail a process whereby new responses become established.

4.10.4: Conclusions

Chapter 4 demonstrated that children with psychopathic tendencies and adult psychopaths are impaired in probabilistic reversal learning. The following chapter will investigate the neural substrates involved in probabilistic reversal learning. This may serve to identify neural regions that might be dysfunctional in children with psychopathic tendencies and adult psychopathic individuals.

Chapter 5 – The Neural substrates of Probabilistic Reversal Learning in Healthy Adults

5.1: Experiment 6

Chapter 4 demonstrated that children with psychopathic tendencies and adult psychopaths are impaired in probabilistic reversal learning. Further, this impairment was demonstrated to be primarily a problem with maintenance of the new response. That is, children with psychopathic tendencies and adult psychopaths were more likely than controls to revert back to the previously correct stimulus after reversing away from it. Experiment 6 aims to further characterize the neural processes involved in probabilistic reversal learning in healthy adults. Doing so may serve to identify neural structures as targets for future research efforts with this population.

Lesion studies with rodents and non-human primates have demonstrated reversal learning impairments following surgical ablation of the OFC (Bohn et al., 2003a; Bohn et al., 2003b; Dias et al., 1996; Dias et al., 1997; Iversen and Mishkin, 1970; Izquierdo et al., 2004; McAlonan and Brown, 2003; Meunier et al., 1997) and striatum (Divac et al., 1967; Taghzouti et al., 1985). This deficit has been attributed to a perseverative tendency (*e.g.* Iversen and Mishkin, 1970). Similarly, neuropsychological work with humans has emphasized the importance of OFC/ventrolateral PFC and striatum (Berlin et al., 2004; Cools et al., 2001; Fellows and Farah, 2003; Hornak et al., 2004; Rahman et al., 1999; Rolls et al., 1994; Swainson et al., 2000). Human neuro-imaging studies of reversal learning have also implicated OFC/ ventrolateral PFC and striatum albeit with some degree of inconsistency across studies (Cools et al., 2002; Kringelbach and Rolls, 2003; Nagahama et al., 2001; O'Doherty et al., 2003a; O'Doherty et al., 2001; Remijnse et al., 2005; Rogers et al., 2000). In addition, there have been reports of activation within dorsomedial frontal cortex during reversal learning (Kringelbach and Rolls, 2003; Nagahama et al., 2001; O'Doherty et al., 2003a; O'Doherty et al., 2001; Remijnse et al., 2005; Rogers et al., 2000).

Despite these studies, however, the functional roles of these regions during reversal learning remain unclear. Previous FMRI work examining reversal learning has associated medial OFC with the representation of reward (O'Doherty et al., 2001) or response maintenance (O'Doherty et al., 2003a). In contrast lateral OFC/ ventrolateral PFC has been associated with the representation of punishment (O'Doherty et al., 2001; Remijnse et al., 2005), the detection of mismatches between expected and actual reinforcement (Kringelbach and Rolls, 2003) or the inhibition of inappropriate behavioural strategies (Cools et al., 2002; Kringelbach and Rolls, 2003; Remijnse et al., 2005). These characterisations led to two contrasting predictions regarding ventrolateral PFC. If it codes punishment information, activation would be expected in response to errors irrespective of phase; alternatively, if it inhibits inappropriate behavioural responses, activation would be expected in response to errors in the reversal phase only.

Reports of striatal activation during reversal learning are compatible with suggestions that the striatum is involved in the prediction/anticipation of reward (Breiter et al., 2001; Knutson et al., 2001; O'Doherty et al., 2003b) and punishment (Seymour et al., 2004). These conceptualisations of striatal activation may be successfully explained by prediction error theory (Cohen, 1997; O'Doherty et al., 2004; O'Doherty et al., 2003b; Seymour et al., 2004). Finally, reports of activation within dorsomedial frontal cortex / dorsal ACC during studies of reversal learning have been attributed to response selection and conflict monitoring (Kringelbach and Rolls, 2003; O'Doherty et al., 2003a; Remijnse et al., 2005).

Typically neuroimaging investigations of reversal learning have employed paradigms whereby the reinforcement contingencies of a single pair of stimuli reverse serially throughout (Cools et al., 2002; Kringelbach and Rolls, 2003; O'Doherty et al., 2003a; O'Doherty et al., 2001; Remijnse et al., 2005). There are two potential shortcomings inherent in this design. Firstly, it is relatively difficult to disentangle acquisition and reversal trials as the acquisition of novel stimuli occurs only once at the beginning of the task. Secondly, it is possible that alternative strategies might be employed (*e.g.*, rule-based strategies such as

"reverse after three incorrect responses"). Indeed, animal lesion studies using this type of 'serially-reversing' paradigm have sometimes reported reversal learning impairment only on the first reversal (Dias et al., 1996; Dias et al., 1997; Iversen and Mishkin, 1970; Schoenbaum et al., 2002).

5.1.1: Summary of Aims

Experiment 5 aimed to use event-related functional magnetic resonance imaging (FMRI) to identify the neural regions involved in reversal learning in healthy humans. In order to minimize the paradigm-limitations associated with previous neuroimaging studies of reversal learning a task was designed (following those used in experiments 4 and 5) whereby participants encountered a number of different stimulus pairs (some of which reversed after a pre-defined number of trials). Following previous FMRI results from human instrumental learning paradigms, it was predicted that medial OFC would be activated by the receipt of reward irrespective of learning phase (acquisition or reversal). Further, experiment 6 allowed a test of two contrasting predictions regarding ventrolateral PFC. If it codes punishment information activation would be expected to errors irrespective of phase; alternatively, if it inhibits inappropriate behavioural responses, activation would be expected in the reversal phase only.

5.2: Methods

5.2.1: Participants

Twenty-eight right-handed adults participated in the study. Due to a technical error, data from one participant was unusable. A further 6 participants were excluded from analysis due to poor performance on the reversal learning task (see methods section *5.2.4.* for further details). Consequently data from twenty-one participants (11 women and 10 men, mean age = 24.71, SD = 2.72; range = 22-34 years) were analysed. All participants were in good health with no past history of psychiatric or neurological disease and gave informed written consent.

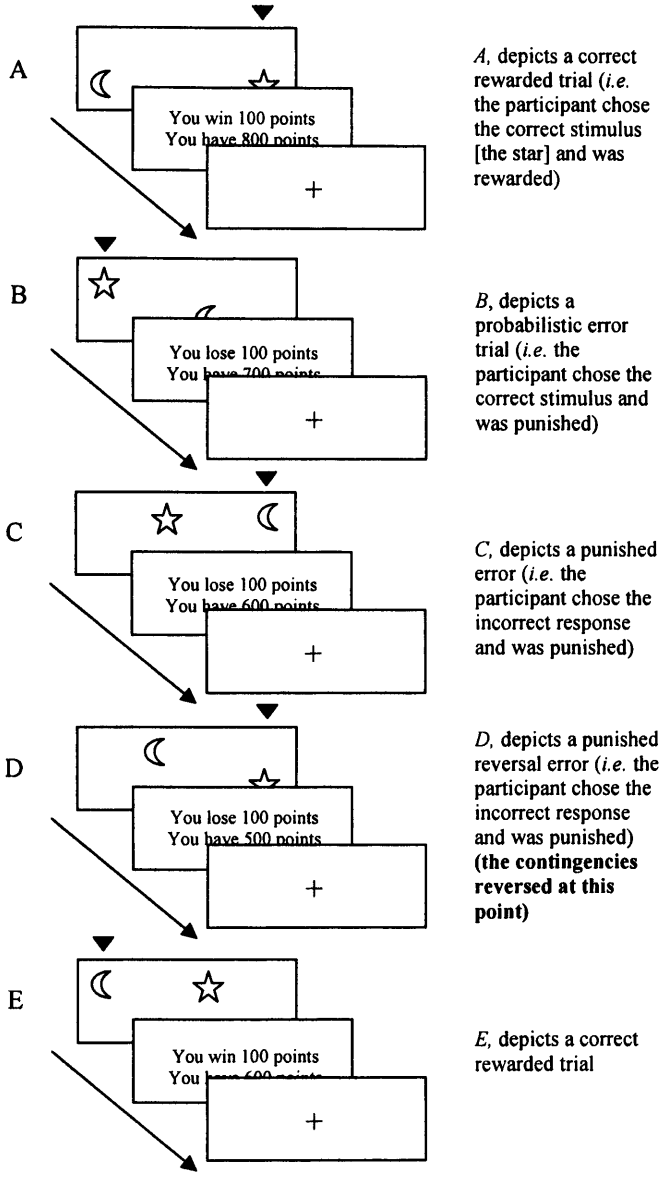
5.2.2: MRI data acquisition

Participants were scanned during task performance using a 3 Tesla GE Signa scanner. In each run a total of 150 functional images were taken with a gradient echo echo-planar imaging (EPI) sequence (repetition time = 2300 ms, echo time = 23 ms, 64 x 64 matrix, flip angle 90°, FOV 24 cm). Whole brain coverage was obtained with 29 axial slices (thickness, 3.3mm; in-plane resolution, 3.75 x 3.75 mm). A high-resolution anatomical scan (three-dimensional Spoiled GRASS; repetition time = 8.1 ms, echo time = 3.2 ms; field of view = 24 cm; flip angle = 20°; 124 axial slices; thickness = 1.0 mm; 256 x 256 matrix) in register with the EPI dataset was obtained covering the whole brain.

3.2.3: Probabilistic Reversal Learning Task and Experimental Procedure

On each trial a pair of stimuli (line drawings of common, neutrally-valenced, items) were displayed on a white background. One stimulus was positioned in one of four possible left-sided screen locations and the other was positioned in one of four possible right-sided screen locations. Participants had to choose a stimulus and received either positive ('you win 100 points') or negative ('you lose 100 points') feedback according to their accuracy and the reinforcement contingency of that pair. Each trial lasted 2300 milliseconds (ms) and involved the presentation of: the test stimuli for 1100 ms, feedback display for 900 ms and finally a fixation cross for 300 ms (see figure *5.1.*). Participants were able to respond by left or right thumb button press (corresponding to selection of the left- or right-positioned stimulus respectively) only during the 1100 ms stimuli presentation window. If participants failed to make a selection they received the following feedback: 'please respond faster next time', and their points total remained the same. Participants began the task with 0 points. A running total of points was visible at the bottom of the screen only during the 900 ms feedback display window.

figure 5.1: The probabilistic reversal learning task



An example of several consecutive trials in the probabilistic reversal learning task is shown (running from top to bottom). On each trial participants are presented with two stimuli (line drawings of common items) for 1100 ms. Using trial-and-error feedback, participants must discover which of the two stimuli is correct (participant choice [right or left stimulus] is indicated by a small arrowhead above the relevant stimulus). Feedback is presented for 900 ms, and then finally, a fixation cross is presented for 300 ms.

Participants were scanned performing the task in eight successive 6 minute runs. Each run comprised three consecutively presented stimulus pairs. This led to a total of 24 pairs, (each of which involved different line drawings). Stimulus pairs were presented for 40 successive trials. The reinforcement contingencies of two of the pairs in each run reversed after 20 trials. The reinforcement contingency of the third pair remained constant throughout. Pairs in half of the runs had a 90-10 probabilistic reinforcement contingency (*i.e.* the participant was rewarded for selecting the correct stimulus on 90% of trials and rewarded for selecting the incorrect stimulus on 10% of trials [the inverse was true for punishment contingencies]), and the other half had a 70-30 probabilistic reinforcement contingency. The order of runs and stimulus pairs within runs was randomised for each participant.

The following instructions, taken from Swainson and colleagues (Swainson et al., 2000), were presented and read aloud to the participant: *'Pairs of objects will appear on the screen. On each go you have to choose one of these objects and the computer will tell you if your choice was correct or wrong. If it is correct you will win 100 points. If it is wrong you will lose 100 points. Each object will sometimes be correct and sometimes be wrong, but one of the objects will tend to be correct more often than the other one. Find out which object is usually correct, and choose that object every time. Stick with it even if it is occasionally wrong. At some point it may change so that the other object is usually correct, in which case you should choose that one every time.'*

In addition to the experimental trials, 24 fixation trials were presented per run to serve as a baseline. Stimuli were presented on a computer display projected onto a mirror in the MRI scanner. The task was programmed in E-Studio. Participants were placed in a light head restraint within the scanner to limit head movement. Prior to the scanning session participants performed a 20 trial probabilistic discrimination training run, without reversal, comprising 1 pair of stimuli. The reinforcement contingency for the practice pair was 80-20.

Trials were split into 8 types according to; the discrimination phase (acquisition/reversal); the participants' response accuracy (correct/incorrect); and

the feedback received (positive/negative). The interaction between response accuracy and feedback was such that participants were able to make one of four responses: (i) a correct response with congruent (*i.e.* positive) feedback, (ii) a correct response with incongruent (*i.e.* negative) feedback, termed a *probabilistic error*, (iii) an incorrect response with congruent feedback, termed a *punished error*, and (iv) an incorrect response with incongruent feedback, termed a *rewarded error*. All trials in the non-reversing conditions were defined as acquisition trials. There were also occasionally trials on which participants did not respond at all, these were modelled as events of no interest.

5.2.4: Behavioural data analysis

Experiment 6 aimed, specifically, to assess successful reversal learning ability, therefore datasets were included in the analysis only if participants met criterion for successful performance (according to a binomial distribution, set at $P < 0.05$). This calculation was performed separately on correct responses (regardless of feedback) for acquisition and reversal trials within each run. If a participant failed to perform at least 2 runs of each contingency successfully according to the binomial distribution (precluding an adequate number of cases of each event) their entire dataset was excluded from analysis. Remaining data were analysed using SPSS. Rates of errors were analysed using a 2 (Contingency; 90-10/70-30) x 2 (Phase; acquisition/reversal) repeated measures ANOVA.

5.2.5: FMRI analysis

Data were analysed within the framework of the general linear model using Analysis of Functional Neuroimages (AFNI) (Cox, 1996). Both individual and group-level analyses were conducted. The first six volumes in each scan series, collected before equilibrium magnetization was reached, were discarded. Motion correction was performed by registering all volumes in the EPI dataset to a volume collected shortly before the high resolution anatomical dataset was acquired. The EPI datasets for each participant were spatially smoothed (using an isotropic 6mm Gaussian kernel) to reduce the influence of anatomical variability

among the individual maps. Next, the time series data were normalized by dividing the signal intensity of a voxel at each time point by the mean signal intensity of that voxel for each run and multiplying the result by 100. Resultant regression coefficients represented a percent signal change from the mean. Following this, regressors depicting each of the 8 response types (and one regressor of no interest modelling the data when participants did not respond) were created by convolving the train of stimulus events with a gamma-variate haemodynamic response function to account for the slow haemodynamic response (Cohen, 1997). The haemodynamic response function was modelled to the onset of the trials (*i.e.* the presentation of the stimuli). Linear regression modelling was performed using the regressors described above plus 2 regressors to model a first order baseline drift function. This produced for each voxel and each regressor, a beta coefficient and its associated t-statistic.

Voxel-wise group analyses involved transforming single-subject beta coefficients into the standard coordinate space of Talairach and Tournoux (Talairach and Tournoux, 1988) and performing a two-sample random effects t-test. Following previous neuroimaging studies of reversal learning (Kringelbach and Rolls, 2003; O'Doherty et al., 2001) punished errors committed in the reversal phase of the task (*i.e. punished reversal errors*) were contrasted with rewarded correct responses committed throughout the task. This resulted in a group map of areas of differential activation ($P < 0.00005$). To correct for multiple comparisons a spatial clustering operation was performed using AlphaSim (Ward, 2000) with 1,000 Monte Carlo simulations taking into account the entire EPI matrix ($P < 0.001$).

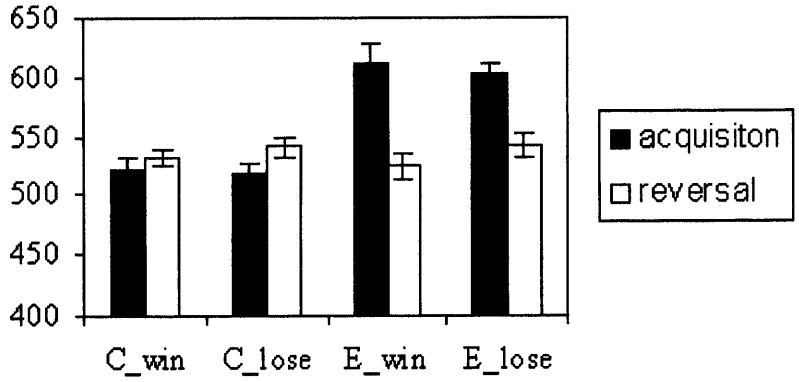
A subset of clusters showing significant differential activation were selected according to *a priori* hypotheses about the regions involved in reversal learning. These clusters were used to define functional regions of interest (ROIs). Finally average percent signal change was measured within each ROI, and these data were analysed using repeated measures analysis of variance (ANOVAs). Mean BOLD responses within these ROIs were examined using a series of 2 (Accuracy) x 2 (Phase) x 2 (Feedback) ANOVAs.

5.3: Results

5.3.1: Behavioural data

Datasets for 6 people were excluded due to poor behavioural performance (see section *5.2.4.*). In the remaining runs, participants made an average 176.67 (s.e. = 5.37) correct responses in the acquisition phases and 208.19 (s.e. = 6.29) in the reversal phases of the pairs that reversed contingency. Reaction time (RT) data was analysed using a 2 (Phase) x 2 (Accuracy) x 2 (Feedback) ANOVA. Significant main effects were revealed for Phase ($F(1,19) = 16.5$, $p < 0.001$) and Accuracy ($F(1,19) = 63.8$, $P < 0.001$). There were also significant interactions between Phase and Accuracy ($F(1,19) = 50.4$, $P < 0.001$), and Phase and Feedback ($F(1,19) = 4.34$, $P < 0.05$). These effects were primarily driven by the increased latencies when performing errors in the acquisition phase (see figure *5.2.*).

figure 5.2: Behavioural results (reaction times)



Reaction Time (RT) data (milliseconds; ms) is shown separately for rewarded correct responses (C_win), probabilistic errors (C_lose), and rewarded (E_win) and punished errors (E_lose). Acquisition and reversal phase responses are represented by black and white bars respectively.

5.3.2: FMRI data

Punished reversal errors were contrasted with all rewarded correct responses. Areas showing significantly greater BOLD responses for punished reversal errors included bilateral ventrolateral PFC, right dorsomedial frontal cortex (extending bilaterally), right middle frontal gyrus, left precentral gyrus, bilateral caudate and bilateral parietal lobe. Areas showing greater activation for rewarded correct responses included right medial OFC (extending bilaterally), bilateral amygdala/ hippocampal regions and left posterior cingulate (see table *5.1.*).

On the basis the predictions outlined in the introduction, five functionally-defined ROIs were identified in: (i) right ventrolateral PFC, (ii) dorsomedial frontal cortex, (iii) right caudate, (iv) medial OFC, and (v) right amygdala/ hippocampus (figures. *5.3.* and *5.4.*). Mean BOLD responses within these ROIs were examined using a series of 2 (Accuracy) x 2 (Phase) x 2 (Feedback) ANOVAs.

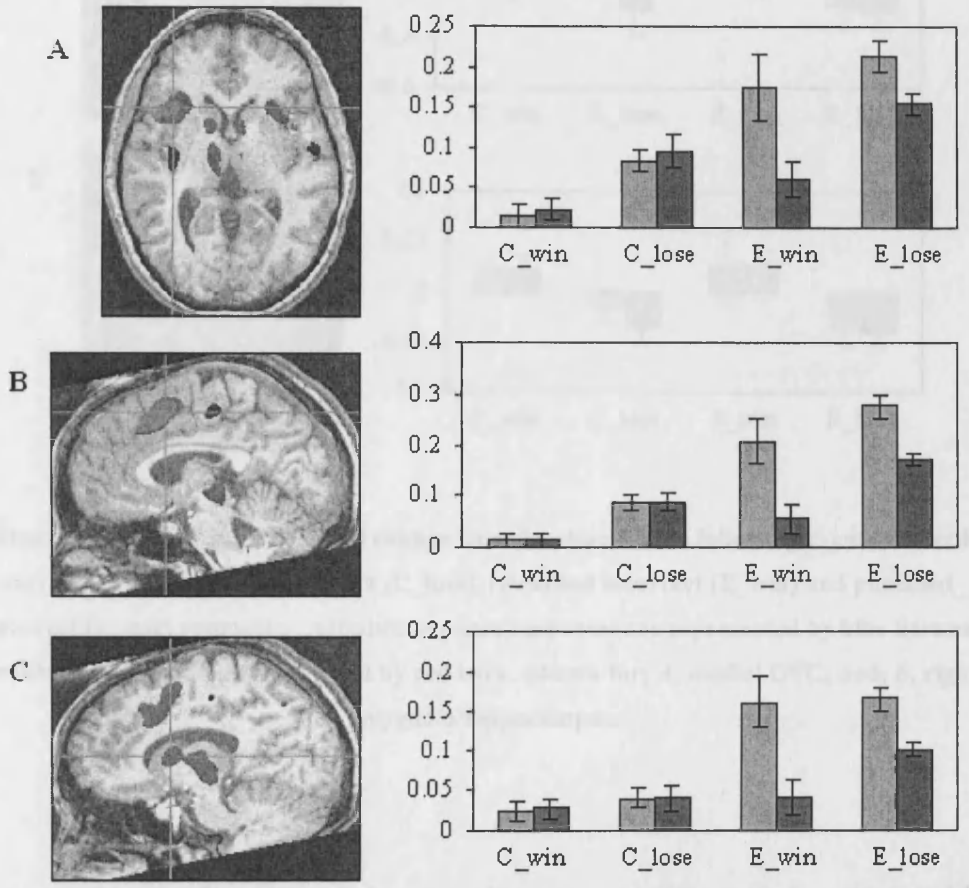
As expected, given the contrast that identified these ROIs, the main effect of feedback was significant for all five ROIs (all $P < 0.001$, except caudate $P < 0.05$). Ventrolateral PFC, dorsomedial frontal cortex and caudate showed significantly greater activity to punishing feedback information (see figure *5.3.*) while medial OFC and amygdala/ hippocampus showed significantly greater activity to rewarding feedback information (see figure *5.4.*). The BOLD response associated with feedback did not interact significantly with those associated with either discrimination phase or accuracy. Significant main effects of phase and accuracy and significant phase by accuracy interactions were revealed for ventrolateral PFC, dorsomedial frontal cortex and caudate (in all cases, P at least < 0.005); BOLD responses within these areas were significantly greater during acquisition and to errors and were particularly strong to errors during acquisition (see figure *5.3.*).

table 5.1: Significant differential activation produced by contrasting punished reversal errors with all rewarded correct responses†

Anatomical Location	l/r	BA	x	y	z	volume(mm^3)
Punished error reversal > Rewarded correct responses:						
Dorsomedial frontal cortex	r	32,6	4	7	53	5907
Middle frontal gyrus	r	9,6,8	52	14	39	4547
Middle frontal gyrus	r	6	33	0	64	253
Precentral gyrus	l	9	-38	24	40	288
Ventrolateral PFC	r	47,13	36	18	7	5108
(extending to anterior insula)						
Insula/ ventrolateral PFC	l	13	-32	15	5	1552
Caudate/thalamus	r		11	6	12	5031
Caudate	l		-7	5	11	267
Inferior parietal lobule	r	40	47	-48	48	1037
Superior/inferior parietal lobule	l	40,7	-30	-59	43	1645
Rewarded correct > Punished error reversal responses:						
Medial OFC*	r	10	3	52	-5	354
Amygdala/ hippocampus	r	34	22	-11	-18	817
Parahippocampal gyrus	l	35,28	-19	-17	-16	2129
(extending to hippocampus/amygdala)						
Posterior cingulate	l	23,30	-1	-55	14	553

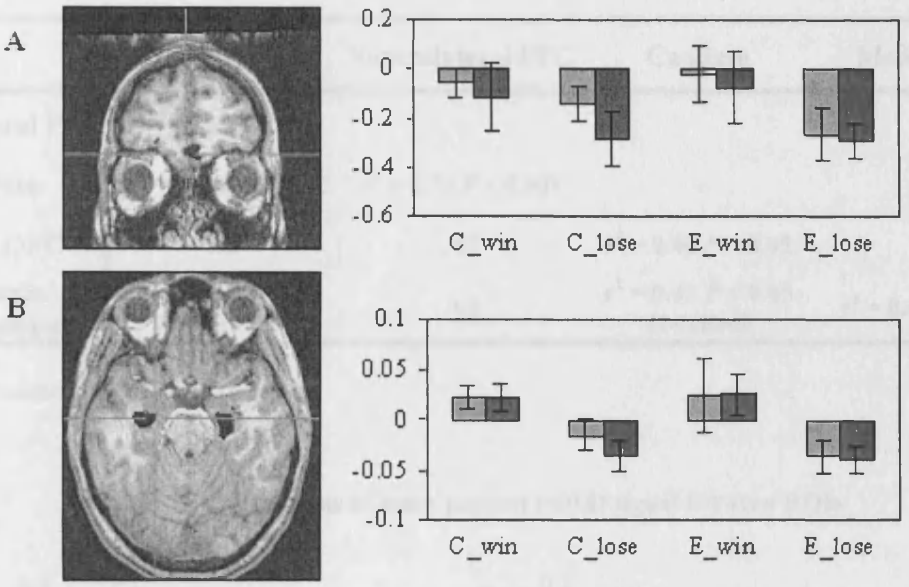
†All activations are significant at $P < 0.001$ (corrected for multiple comparisons), except * significant at $P < 0.00005$ uncorrected for multiple comparisons.

figure 5.3: Areas showing greater activation to punished reversal errors



Histograms detailing percent signal change from baseline for the following events: rewarded correct (C_win), punished correct (C_lose), rewarded incorrect (E_win) and punished_ incorrect (E_lose) responses. Acquisition phase responses are represented by blue bars and reversal responses are represented by red bars. Shown for; *A*, right ventrolateral PFC; *B*, dorsomedial frontal cortex; and, *C*, right caudate.

figure 5.4: Areas showing greater activation to rewarded correct responses



Histograms detailing percent signal change from baseline for the following events: rewarded correct (C_win), punished correct (C_lose), rewarded incorrect (E_win) and punished_ incorrect (E_lose) responses. Acquisition phase responses are represented by blue bars and reversal responses are represented by red bars. Shown for; *A*, medial OFC; and, *B*, right amygdala/hippocampus.

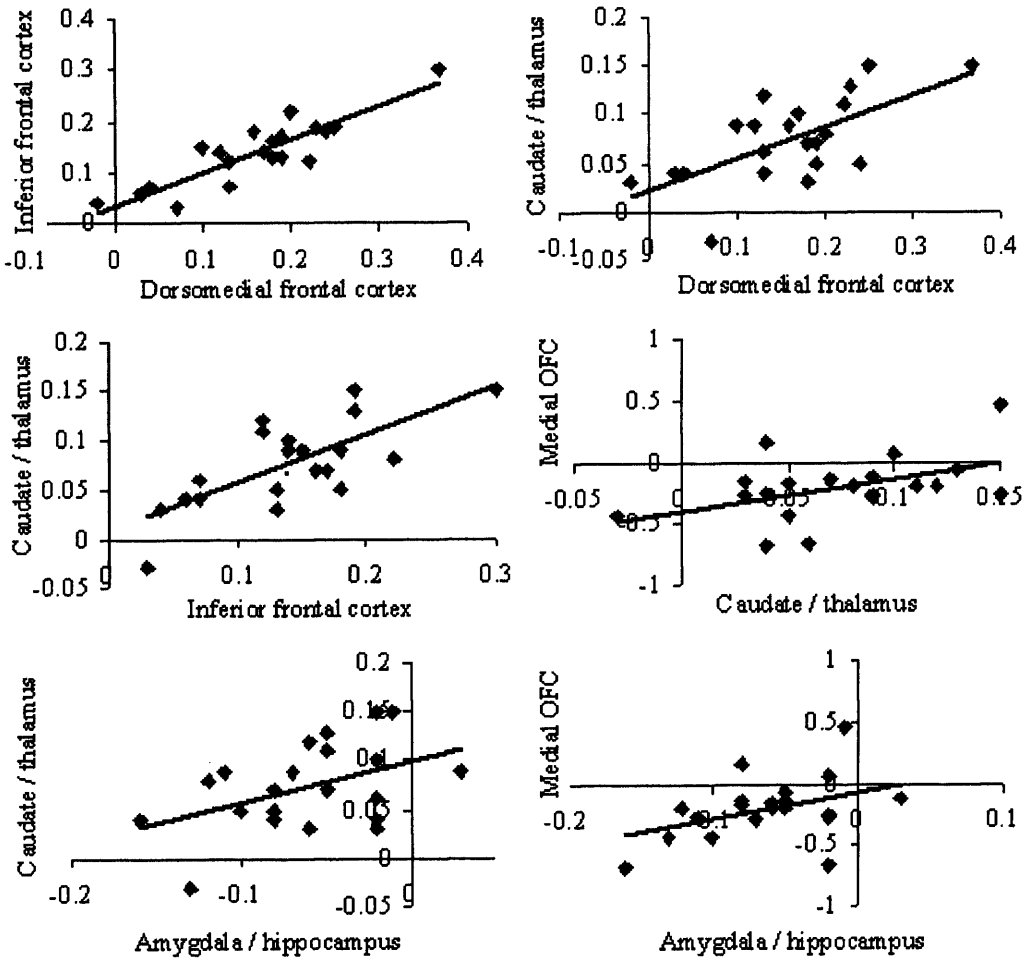
Subsequently, a correlational analysis was used to examine relationships between differences in activity, within each ROI, with respect to punished reversal errors and rewarded correct responses. These revealed highly significant inter-relationships between activity changes measured in ventrolateral PFC, dorsomedial frontal cortex and caudate (see table *5.2.* & figure *5.5.*). In other words, within these regions, the increase in BOLD response during punished reversal errors relative to rewarded correct responses were highly correlated. There were also relationships between activity changes across conditions in caudate, medial OFC and amygdala/ hippocampus, however these correlations were somewhat lower (see table *5.2.* & figure *5.5.*).

table 5.2: Significant correlations between each ROI in the difference in BOLD signal between punished reversal errors and all rewarded correct responses

	Dorsomedial Frontal Cortex	Ventrolateral PFC	Caudate	Medial OFC
Ventrolateral PFC	$r^2 = 0.86$ $P < 0.001$	-		
Caudate	$r^2 = 0.65$ $P < 0.0015$	$r^2 = 0.75$ $P < 0.001$	-	
Medial OFC	*NS*	*NS*	$r^2 = 0.46$ $P < 0.05$	-
Amygdala/ Hippocampus	*NS*	*NS*	$r^2 = 0.42$ $P < 0.05$ (1-tailed)	$r^2 = 0.43$ $P < 0.05$

NS = non-significant correlations.

figure 5.5: Correlations of mean percent BOLD signal between ROIs



Values represent the difference in BOLD activation between punished reversal errors and rewarded correct responses made throughout the task.

5. 4: Discussion

Experiment 6 used event-related FMRI to examine BOLD responses associated with performance of a probabilistic reversal learning task. Punished reversal errors were associated with significantly greater activity than rewarded correct responses in ventrolateral PFC, dorsomedial frontal cortex and caudate. Investigation of the BOLD signal changes associated with specific events within these ROIs revealed, unexpectedly, that the error-related activity occurred during both acquisition and reversal. Indeed, the BOLD responses were generally greater for errors during acquisition. Punishment, relative to reward, was associated with increases in activation in ventrolateral PFC, dorsomedial frontal cortex and caudate but decreases in activation in medial OFC and amygdala/ hippocampus. Correlational analyses indicated two influences on behaviour as indexed by caudate activity: one involving inferior and dorsomedial frontal cortex and the second involving medial OFC and the amygdala/ hippocampus.

Previous neuroimaging studies of reversal learning have associated lateral OFC/ ventrolateral PFC with the representation of punishment (O'Doherty et al., 2001; Remijnse et al., 2005) or the suppression of previously rewarded responses (Cools et al., 2002; Elliott et al., 2000a; Monchi et al., 2001). These positions led to divergent predictions regarding the current data. If this area codes punishment information activation should have been expected during errors in both the acquisition and reversal phases, however if it inhibits previously appropriate behavioural responses then activation should have expected in response to errors during the reversal phase only. The current data was in line with the former hypothesis; ventrolateral PFC was activated on punished trials not only in the reversal phase, but also during acquisition; *i.e.*, before response suppression was necessary. Moreover, and problematic for both positions, ventrolateral PFC also showed significant activation to *rewarded* errors during acquisition. Thus, on the basis of these results, it appears that ventrolateral PFC may be coding for accuracy rather than punishment *per se*.

Highly significant inter correlations were observed between the increase in BOLD response for punished reversal errors relative to rewarded correct

responses in inferior and dorsomedial frontal cortex and caudate. Activation within these regions has been reported in previous neuroimaging studies of reversal learning (Cools et al., 2001; Kringelbach and Rolls, 2003; Nagahama et al., 2001; O'Doherty et al., 2003a; O'Doherty et al., 2001; Remijnse et al., 2005; Rogers et al., 2000; Swainson et al., 2000). In particular the results are similar to those of Kringelbach and Rolls who observed similar patterns of activity within lateral OFC and dorsal ACC (Kringelbach and Rolls, 2003). These data are also in line with previous human neuropsychological work which has emphasized the role of striatal regions in reversal learning (Cools et al., 2001; Swainson et al., 2000). In experiment 6 ventrolateral PFC, dorsomedial frontal cortex and caudate all showed significantly greater response to punishment than reward during both phases, and significantly greater responses to errors rather than correct responses. Interestingly, this error-related activity was irrespective of feedback during the acquisition phase, that is, it occurred even when errors were rewarded. Moreover, within these regions, the degree of increase in BOLD response to punished reversal errors relative to rewarded correct responses were highly correlated across participants (see figure *5.5.*). On the basis of the results from experiment 6, it appears that an integrated response is executed within these regions to modulate behaviour during reversal learning.

It is commonly suggested that dorsal anterior cingulate cortex (ACC), a region proximal to the dorsomedial frontal cortical activation observed here, is implicated in monitoring response conflict (Botvinick et al., 1999; Botvinick et al., 2004; Bush et al., 2000; Carter et al., 1998; Kerns et al., 2004; Kringelbach and Rolls, 2003; Schall et al., 2002). Further, activation of dorsal ACC has often co-occurred with activation in dorsolateral PFC (DLPFC) especially during performance of attentional tasks (Botvinick et al., 1999; Kerns et al., 2004). (Botvinick et al., 1999; Kerns et al., 2004). In keeping with this, and consistent with a recent study by Remijnse and colleagues (Remijnse et al., 2005) activation was also observed in DLPFC (BA 9) during commission of punished reversal errors. Concurrent activation within these regions following response conflict has been used as evidence indicating that ACC plays a role in augmenting a

representation of stimulus features within DLPFC. Specifically, it has been suggested, that this augmentation results in increased cognitive/ attentional control of stimuli represented in temporal cortex (Garavan et al., 2002; MacDonald et al., 2000; Ruff et al., 2001). On the basis of the results of experiment 6 it appears that an analogous account may be used to explain the relationship between dorsomedial and ventrolateral PFC. Specifically, that dorsomedial frontal cortex, in situations of response conflict, augments the representation of object/ motor features within ventrolateral PFC that allow control over *motor* responding, which is mediated by the caudate. Importantly, the functions outlined above would not be expected to occur only in reversal. Activation of these regions ought to be a consequence of response conflict also during acquisition of a stimulus-response association. In light of this explanation, it is less surprising that these systems are activated by punishment signals; punishment indicates an erroneous response and should therefore engender response conflict. It is also less remarkable that activation was also observed in these regions to *rewarded* errors. The commission of errors ought to be associated with response conflict, regardless of feedback. In line with this position, ACC/ dorsomedial frontal cortex, inferior / ventrolateral frontal cortex and caudate activation has been reported in other paradigms that evoke motor response conflict (Bush et al., 1999; Casey et al., 2002; Casey et al., 2000; Menon et al., 2001; Peterson et al., 2002).

Importantly, this account suggests that activation of dorsomedial and ventrolateral PFC ought to be a consequence of response conflict during acquisition and reversal of stimulus-response associations, this circuit is not the only region implicated in the acquisition of a stimulus-response association. Indeed, neuropsychological data indicate that, following lesions to the ventrolateral PFC, the ability to acquire new stimulus-response discriminations remains intact (Fellows and Farah, 2003; Hornak et al., 2004; Rolls et al., 1994). Animal data indicate that stimulus-response associations may be acquired through the interaction of temporal cortical regions and caudate (Messinger et al., 2001; Packard, 1999; Packard and McGaugh, 1996). The augmentation of a motor response need not be necessary for response acquisition. In contrast, however, the

augmentation of a competing motor response would be particularly important in reversal; the newly incorrect stimulus-response habit would need to be overridden and superseded by the newly correct response. Indeed, behavioural data suggest that successful reversal learning does reflect acquisition of a competing response (Rescorla, 1996).

Activation within medial OFC has previously been associated with the representation of reward (Elliott et al., 2000a; Kringelbach and Rolls, 2004; O'Doherty et al., 2001) or response maintenance (O'Doherty et al., 2003a). In keeping with these observations it is interesting to note that the activation in medial OFC observed in experiment 6 was not modulated by accuracy or discrimination phase – medial OFC activity here was modulated exclusively by feedback. Further, consistent with the results of O'Doherty and colleagues (O'Doherty et al., 2003a) the differences in percent signal change reflected reduced deactivation following the receipt of reward in contrast with punishment. Following Schoenbaum and colleagues, these data are consistent with the notion that medial OFC codes the expectation of reinforcement, which is acquired via input from amygdala, and is used to guide behaviour (Schoenbaum et al., 2003). Specifically, that the reception of punishment leads to disruption of the reinforcement expectation signal guiding behaviour and perhaps signals that a change in behaviour is required. Consistent with this idea, it is interesting to note the positive correlations between medial OFC, caudate and amygdala/hippocampus.

5.4.1: Summary and Conclusions

Experiment 6 investigated reversal learning under conditions of probabilistic feedback. Errors were associated with increased activation in a circuit including right ventrolateral PFC, dorsomedial frontal cortex and right caudate. In contrast, reward, independent of accuracy, was associated with activation in medial OFC and amygdala/hippocampus.

5.5: General Discussion

As discussed in the previous chapter (section *4.10.2.*) it appears that there exits, in individuals with psychopathy, an inability to successfully override a previous motor response habit and/or form a stable new motor response habit. The results of experiment 6 lend further support to the suggestion that the neural basis of the reversal learning impairment is within ventrolateral PFC.

These data provide a strong impetus for a modification to the integrated emotion systems (IES) model (Blair, 2004). Specifically, that ventrolateral PFC was responsive to rewarded errors, and also during the acquisition phase indicates that this region is not specifically involved in gating previously adaptive motor responses. Instead, and consistent with the results of experiments 4 and 5, the results of experiment 6 indicate that ventrolateral PFC is important in boosting representations, and particularly in overriding previously correct motor responses. Further, and in line with suggestions made in section *3.5.* activation within medial OFC is consistent with the idea that this area codes for accuracy and thus may be a critical substrate in signalling if expectation violations have occurred.

5.6: Conclusions

In conclusion, experiment 6 investigated the neural substrates involved in successful probabilistic reversal learning. In line with predictions, and previous literature, a circuit including ventrolateral PFC, medial OFC, dorsomedial frontal cortex, caudate and amygdala was activated. These results provide an interesting interpretation of the reversal learning deficit present in individuals with psychopathy and provide a potential refinement for the IES model of emotional learning.

Chapter 6 – Discussion and Future Directions

6.1: Summary of Experimental Work

This thesis began by introducing the phenomenon of psychopathy, a developmental disorder that is associated with an increased risk for violence and criminality in child- and adulthood. Also, the important issue of co-morbidity between psychopathic tendencies and attention-deficit hyperactivity disorder (ADHD) in childhood was introduced. The second half of Chapter 1 went on to discuss and evaluate six theories attempting to explain psychopathy. Particularly pertinent to this thesis, it was noted that the fear (*e.g.* Cleckley, 1976; Eysenck, 1964; Fowles, 1988; Gray, 1987; Lykken, 1995; Mealey, 1995; Patrick, 1994; Pichot, 1978; Trasler, 1978; Trasler, 1973), response-set modulation (RM; Newman, 1998; Patterson and Newman, 1993) and integrated emotion systems (IES; Blair, 2004) models all provided different accounts of the passive avoidance and reversal learning deficits observed in psychopathic individuals. It was also noted, however, that the nature of these deficits in children with psychopathic tendencies is as yet unclear. Indeed, the remainder of the thesis was dedicated to the investigation of passive avoidance learning and reversal learning. Firstly, attempts were made to characterise the deficits in children with psychopathic tendencies and then to go onto investigate the neural substrates involved in these processes during performance by healthy adults.

Chapter 2 reviewed previous experimental data collected using the passive avoidance learning task with adult psychopaths and children with psychopathic tendencies which indicate that whilst adults with psychopathy have consistently presented with impairment in passive avoidance learning, children with psychopathic tendencies have not. It was also noted that previous research has indicated that passive avoidance learning may be impaired in children with ADHD. Experiment 1 assessed passive avoidance learning in children with psychopathic tendencies and removed any variation due to level of ADHD. In

addition, a paradigm modification was introduced to determine whether the two groups would be differentially affected by varying the level of reward/ punishment associated with individual stimuli. The results of experiment 1 replicated previous findings with psychopathic adults indicating that psychopathic tendencies in children are associated with poor passive avoidance learning. Further, passive avoidance error rates were modulated in comparison children by modifying the incentive value of the CS-s. This effect was not observed in children with psychopathic tendencies. Experiment 2 developed a Hebbian Learning model of passive avoidance learning in order to further characterize the impairment shown by children with psychopathic tendencies in experiment 1. In order to simulate the various predictions the model was either left intact, impaired with regard to all stimulus-reinforcement learning, or selectively impaired with regard to stimulus-punishment learning. The model was found to capture the performance of the comparison children very successfully in terms of both passive avoidance and omission error data. In contrast, the data of the children with psychopathic tendencies was most successfully captured by a specific impairment in the formation of stimulus-punishment associations.

Chapter 3 began with a review of previous animal and human instrumental learning literature, concluding that a network including medial OFC/ rostral ACC, the insula, striatum, hippocampus and amygdala had been identified. Experiment 3 assessed the BOLD responses associated with passive avoidance learning in healthy human participants, specifically late correct responses were contrasted with early correct responses and incorrect responses. The results of experiment 3 revealed, consistent with the animal literature, that successful passive avoidance learning requires the appropriate recruitment of regions including rostral ACC, insula, caudate, hippocampal regions, and the amygdala.

Chapter 4 introduced the phenomenon of reversal learning and reviewed previous experimental data collected with adult psychopaths and children with psychopathic tendencies, concluding that whilst adults with psychopathy have consistently presented with impairment in reversal learning, children with psychopathic tendencies have not. It was also noted that a previous report

indicated that reversal learning may be impaired in children with ADHD. It was suggested that, on the basis of the previous literature, a critical factor in the success of children with psychopathic tendencies may the salience of the contingency change. In order to test this hypothesis experiment 4 assessed the performance of children with psychopathic tendencies on a novel probabilistic reversal learning paradigm with four different contingencies (and removed any variation due to level of ADHD). The results of experiment 4 revealed that children with psychopathic tendencies presented with impairment only on the probabilistic contingencies. Moreover, as contingencies became more probabilistic the impairment became more pronounced. Further, it was found that the children with psychopathic tendencies committed more win-shift errors only in the reversal phases. It was suggested that these data replicated previous investigations of reversal learning and extinction in children with psychopathic tendencies. It was also suggested that they lend further support to the idea that the salience of contingency change may be a critical factor in the reversal learning ability of children with psychopathic tendencies.

Experiment 5 investigated probabilistic reversal learning ability, as a function of salience of contingency change, in a group of adult individuals with psychopathy. Given the results obtained in experiment 4, experiment 5 aimed to further characterise the impairment in adult psychopaths, and particularly to examine the data using a win-stay, lose-shift analysis. In line with predictions, results of experiment 5 revealed that adults with psychopathy showed a performance deficit in both the simple and probabilistic condition relative to comparison individuals. Also, similarly to the children, the adults with psychopathy committed more win-shift responses, specifically in the reversal phases. It was suggested that these data are line with previous suggestions that the reversal learning impairment in adult psychopaths is more pronounced than that in children with psychopathic tendencies. Potential reasons for this difference were discussed.

Chapter 5 reviewed the previous animal and human literature investigating the neural substrates associated with reversal learning, concluding that

consistently medial/orbital and ventrolateral regions, in addition to dorsal ACC and striatum have been identified. Experiment 6 aimed to investigate the neural substrates involved in successful probabilistic reversal learning with healthy participants, using a novel task which attempted to remediate some shortcomings present in previous investigations. Following previous studies, punished errors made in the reversal phase were contrasted with rewarded correct responses made throughout the task. In line with predictions, the results of experiment 6 revealed that punished reversal errors were associated with BOLD increases in dorsomedial and ventrolateral PFC and caudate. Surprisingly, investigation of the percent signal change for different events within these regions revealed that these areas were also active during punished errors in the acquisition phase. Punished reversal errors were also associated with BOLD decreases in medial OFC and amygdala/hippocampus.

6.2: Implications for the Theories of Psychopathy

The experimental research presented in this thesis has implications for several theories attempting to explain psychopathy. This section will go on to summarize implications for four of the theories presented in section *1.3*.

6.2.1: The Frontal Lobe Dysfunction Hypothesis

The FL hypothesis (Barratt, 1994; Elliot, 1978; Gorenstein, 1982; Moffitt, 1993a; Raine, 1997; Raine, 2002a) attempts to describe the association between frontal damage, antisocial behaviour and psychopathy. This position was found to successfully *describe* the association between frontal dysfunction and aggression. As an account of psychopathy, however, it is unsuccessful. Specifically, it is able to describe the reactive aggression associated with psychopathy which, it appears, may be a result of OFC/ ventral; PFC dysfunction. This position, however, is unable to describe the 'core' instrumental aggression that is relatively unique to this disorder. Indeed, it is unclear how this position would be able to differentiate between the different types of aggression displayed by individuals with psychopathy. Further, a general 'frontal lobe' impairment position would probably predict wide-ranging executive dysfunction, which is not generally observed in psychopathy (LaPierre et al., 1995; Roussy and Toupin, 2000). In conclusion, this neural-level theory requires further specification at the cognitive level. In particular, abnormal functioning of the ventral PFC does not sufficiently account for the range of data available regarding individuals with psychopathy beyond the presence of reversal learning deficits.

6.2.2: The Fear Dysfunction Hypotheses

The fear positions (*e.g.* Cleckley, 1976; Eysenck, 1964; Fowles, 1988; Gray, 1987; Lykken, 1995; Mealey, 1995; Patrick, 1994; Pichot, 1978; Trasler, 1978; Trasler, 1973) would successfully predict impaired passive avoidance learning performance, owing to an insensitivity to punishment. However, it must be noted that the pattern of performance on the *rewarded* trials was also abnormal

in children with psychopathic tendencies. In contrast, the pattern of performance of the comparison children was predicted relatively successfully by the connectionist model, both in terms of reward- and punishment-related processing; in both cases there was a linear relationship between reinforcement and error rate. In contrast, the pattern of performance of the children with psychopathic tendencies was only marginally successfully predicted by the specific stimulus-punishment impairment model; the specific stimulus-punishment account did not predict the abnormal pattern of performance as regards omission errors. In short, it appears that the ability to modulate performance as a function of punishment, but also to a lesser degree, reward, is impaired in children with psychopathic tendencies. The fear positions explicitly predict impaired passive avoidance learning due to an insensitivity to punishment information, as such they are unable to account for overall performance.

The reversal learning data presented in chapter 4 is also problematic for the fear positions. Firstly, these models are unable to account for the dissociation between performance in acquisition and reversal phases. Essentially, it is unclear how the model would explain why individuals with psychopathy are only insensitive to punishment in the reversal phase. Further, and even more problematic for these positions, the results from the win-stay, lose-shift analysis suggest that rather than perseverating with a punished response, as would be predicted by these positions, both children with psychopathic tendencies and adult psychopaths perform increased rates of win-shift responses. In short, after successfully reversing responses away from the newly incorrect (and newly punished) stimulus, the individuals with psychopathy reverse their responses back towards the newly incorrect stimulus. Thus, it appears that individuals with psychopathy are able to process punishment information in some cases (as indexed by their lose-shift responses).

6.2.3: The Response-Set Modulation hypothesis

The RM model (Patterson & Newman 1993; Newman 1998) predicts that individuals with psychopathy will be more likely than comparison individuals to

respond even when the salient stimulus (*i.e.* a CS-) is present on the screen in tasks of passive avoidance learning. As regards the RM hypothesis, similar criticisms as were suggested above apply. Namely, that there was evidence of an abnormal pattern of performance in the reward-related trials, thus an explanation focusing on punishment is essentially inadequate. Further, according to the model, *"the impulsivity, poor passive avoidance, and emotion-processing deficits of individuals with psychopathy may all be understood as a failure to process the meaning of information that is peripheral or incidental to their deliberate focus of attention"* (Lorenz & Newman, 2002; p. 92). However, and as noted in section *1.3.5.*, both the passive avoidance learning task and the reversal learning tasks used in this thesis, presented the punishment information appears on screen *in the absence of any competing information.* Thus it is unclear how this attentional explanation can account for the deficit.

Similarly, the reversal learning data presented in chapter 4 are also problematic for the RM hypothesis. Particularly, the RM hypothesis appears unable to account for the dissociation between performance in acquisition and reversal phases, and also the propensity to perform win-shift (rather than lose-stay responses). Essentially, it is unclear how the model would predict that the inability to divert attention from the goal of gaining reward to 'peripheral' punishment information would be present only in the reversal phase. Further, it appears that individuals with psychopathy are able to attend to punishment information (as indexed by normal rates of lose-shift responses) in the reversal learning paradigm.

6.2.4: The Integrated Emotion Systems Model

Finally the IES model (Blair, 2004) was able to account for much of the data presented in this thesis. Importantly, the conceptualisation of fundamental differences between the neural substrates involved in stimulus-reinforcement and stimulus-response learning were supported. The model would predict impairment in the formation of stimulus-reinforcement associations as indexed by the passive

avoidance learning data in chapter 2. Indeed, it also specifies that whilst the ability to form stimulus-reinforcement associations is impaired, the formation of stimulus-punishment associations is impaired to a greater degree than stimulus-reward associations. This would successfully account for abnormal patterns of passive avoidance and omission errors.

Further, the IES model predicts that the ability of individuals with psychopathy to perform stimulus-response associations remains intact, whilst the ability to *reverse* these stimulus-response associations is impaired. This would allow for the observed dissociation between performance in acquisition and reversal phases that was problematic for the fear and RM hypotheses. The results of the win-stay, lose-shift analysis, and experiment 6, indicate, however, that the IES model requires modification. Specifically, the IES model states that ventrolateral PFC comprises units involved detecting when reinforcement expectations have been violated and in gating previously appropriate (and now inappropriate) motor responses. This is hypothesized to be implemented by means of altering the strength of connections between amygdala and medial OFC. Instead, on the basis of data in chapters 4 and 5, it appears that the ventrolateral PFC acts to allow new responses to become established. Further, the role ascribed to ventrolateral PFC of the detection of expectation violations appears to be fulfilled rather by medial OFC.

6.3: Implications of these Results for the Characterization of Psychopathy and Possible Causes of these Deficits

In line with the IES model, the research presented in this thesis suggests that psychopathy is caused by an impairment in the performance of specific forms of emotional learning, namely stimulus-reinforcement, (exemplified by passive avoidance learning), and reversal learning. The impairments in passive avoidance and reversal learning in adult psychopaths and children with psychopathic tendencies may have implications regarding the presentation of the disorder. This section will refer back to the review of behavioural observations and symptoms presented in section *1.2.2.* and will attempt to integrate these with the experimental results obtained in this thesis.

6.3.1: Passive Avoidance Learning and Empathy

According to the IES model of emotional learning, psychopathy is associated with amygdala dysfunction (Blair, 2004; Blair, 2001; Blair, 2003b; Blair et al., 1999; Patrick, 1994). This dysfunction manifests, at the cognitive level as, an impaired ability to associate a stimulus with an affect representation (*i.e.* to perform passive avoidance learning). Importantly, it appears that this dysfunction may form the core of the deficit in psychopathy; that is, the impaired ability to form stimulus-reinforcement associations could lead to the presentation of *CU* traits. Essentially, it appears that the emotional deficit interferes with socialization such that individuals with the disorder do not find the prospect of goal directed antisocial behaviour aversive (importantly, this does not suggest that the emotional deficit associated with psychopathy in itself motivates the individual to offend). Such an explanation would account for empirical data indicating that individuals with psychopathy display high levels of instrumental aggression, possess anomalous concepts of guilt, are shallow and manipulative, display impaired recognition of fearful expressions, demonstrate attenuated aversive conditioning and startle reflex modulation, and, of course, impaired passive avoidance learning (*e.g.* Aniskiewicz, 1979; Blair, 2001a; Blair et al.,

2001c; Blair, 2003a; Blair, 1995; Blair, 1997; Blair et al., 1995a; Blair, 1999; Cornell et al., 1996; Flor et al., 2002; Frick, 1995; Frick et al., 2000; Hare, 1991; Hare and Quinn, 1971; Harpur et al., 1988; Harpur et al., 1989; House and Milligan, 1976; Kosson et al., 1990; Lang et al., 1990; Levenston et al., 2000; Lykken, 1957; Newman and Kosson, 1986; Newman et al., 1990; Newman and Schmitt, 1998; Pastor et al., 2003; Patrick, 1994; Sutker, 1970; Thornquist and Zuckerman, 1995; Williamson et al., 1987; Woodworth and Porter, 2002). Interestingly, many (though not all) of these impairments are also presented by individuals with amygdala damage (*e.g.* Angrilli et al. 1996; Bechara et al. 1995; Fine, 2000; Fine and Blair, 2000; LaBar et al. 1995), lending further support to the hypothesis that there exits amygdala dysfunction in psychopathy (Blair, 2001; Blair, 2003a; Blair, 2004; Patrick, 1994).

It has been suggested that genetic abnormalities may give rise to the purported amygdala dysfunction, specifically, that the abnormality may manifest as a specific deficit in neurotransmitter function (Blair, 2004). In keeping with this idea, as noted in section *1.2.2.* strong evidence has recently been reported for a substantial genetic component to *CU* traits (Viding et al., 2005). However, it remains unclear which neurotransmitter systems might be dysfunctional in individuals with psychopathy. A candidate would be the noradrenergic system. In line with this suggestion, McIntyre and colleagues (McIntyre et al., 2002) reported that, in rodents, noradrenaline levels within amygdala predicted prior retention of passive avoidance associations. Also, it has been recently suggested that that noradrenaline is involved in mediating the impact of aversive cues in human choice (Rogers et al., 2004a). Moreover, recent pharmacological data imply that noradrenergic manipulations selectively impact on the processing of sad expressions (Harmer et al., 2001; Sustrik et al., in preparation). Thus, a possibility that requires further study is that the putative genetic anomalies disrupt the functioning of the noradrenergic system such that the impact of aversive stimuli is muted.

6.3.2: Reversal Learning and Reactive Aggression

Experiment 6 (and other neuroimaging and neuropsychological studies) have indicated a clear role of ventrolateral frontal cortex in reversal learning. Further, on the basis of the results of experiment 6, it was suggested that ventrolateral PFC may be important in 'boosting' a representation of competing motor responses in situations where response conflict has been signalled (thus suppressing the activation of the prepotent response). In essence it may be suggested that ventrolateral PFC is involved in the regulation of on-line instrumental behaviour; particularly with respect to changing this behaviour following changes in contingency or task demands. Further, the results of experiments 4 and 5 suggested that this function may be impaired in individuals with psychopathy, leading to an increased tendency to shift away from a correct, rewarded response and return to an incorrect, punished response (the previously correct response). Thus, it may be tentatively inferred on the basis of the pattern of deficits present in experiments 4 and 5, in conjunction with the results of experiment 6, suggest that there exists dysfunction within ventrolateral PFC in psychopathy.

It was suggested in section *4.10.2* that the underlying dysfunction that renders individuals with psychopathy unable to engage efficiently in reversal learning may be the same dysfunction that causes these individuals to engage in reactive aggression. Indeed, a further commonality between individuals with psychopathy and individuals sustaining lesions to orbital, ventromedial, or ventrolateral frontal cortex, in addition to reversal learning impairment, is the increased propensity to engage in reactive aggression (Anderson et al., 1999; Barrash et al., 2000; Cornell et al., 1996; Grafman et al., 1996; Pennington and Bennetto, 1993; Williamson et al., 1987; Woodworth and Porter, 2002). Moreover, Rolls and colleagues (Rolls et al., 1994) noted that impairment in reversal learning ability correlated with socially inappropriate and 'disinhibited' behaviour. A dedicated neural circuitry allows the expression of reactive aggression in mammalian species, with the final tier being explosive attack behaviour (Blanchard et al., 1977). This basic threat circuitry is regulated by

executive systems which are considered to be able to either augment or suppress the baseline level of stimulation of the basic threat circuitry. Following from this, it has been suggested that this regulation may be executed by medial, orbital and inferior frontal regions (Anderson et al., 1999; Grafman et al., 1996; Gregg and Siegel, 2001; Panksepp, 1998; Pennington and Bennetto, 1993). The increased propensity towards the display of reactive aggression may be essentially due to a reduced ability to reverse responses resulting in an increased tendency to become frustrated. Indeed, it is well known that frustration is a cue for aggression (Berkowitz, 1993). In addition, ventrolateral PFC has been activated in neuroimaging studies when individuals were induced to feel angry (Dougherty et al., 1999) and also when processing situations that were likely to cause anger (Berthoz et al., 2002).

In terms of the mechanism of dysfunction, the 5-hydroxytryptamine (5-HT) may be a candidate neurotransmitter system. In a reversal learning paradigm rodents injected with a neurotoxin selective for 5-HT were impaired (Clarke et al., 2004; Clarke et al., 2005). Moreover, reversal learning has been impaired, by tryptophan depletion, in some studies in humans (Rogers et al., 1999a), whilst in others has just led to slowing effects (Murphy et al., 2002). Further, studies have reported a negative correlation between criminality/reactive aggression and 5-HT function (Alm et al., 1996; Dolan and Anderson, 2003; Goveas et al., 2004). Indeed, 5-HT has long been implicated in the modulation of (particularly reactive) aggression and impulsivity (Alm et al., 1996; Dolan and Anderson, 2003; Goveas et al., 2004). Specific gene knock-out studies on mice have reported increased aggressiveness for several knock-outs affecting serotonergic functioning including the $5\text{-}HT_{1B}$ receptor (Ramboz et al., 1996) and MAO (monoamine oxidase) A but not B (Shih et al, 1999). Interestingly, recent work has suggested the possibility that the emergence of aggression might be a necessary interaction of environmental stressors with particular genetic contributions to the functioning of the serotonergic system (Moffitt et al., 2002). Thus, Caspi et al (2002) observed that a functional polymorphism in the gene encoding MAOA moderated the effect of maltreatment. Maltreated children with a genotype conferring high levels of

MAOA expression were less likely to develop antisocial problems than maltreated children with a genotype conferring low levels of MAOA expression. Thus, a possibility that requires further study is that the putative dysfunction within ventrolateral PFC leads to reduced 5-HT transmission within this area.

6.5: Limitations of Current Work and Future Directions

There were some limitations of the experimental work presented in this thesis. Firstly, all individuals included in the sample were male. Research which assesses the cognitive impairments present in female psychopaths would be valuable (Cale and Lilienfeld, 2002). Indeed, it has been suggested that there may exist gender differences in certain neurocognitive tasks, especially those relying upon orbital and ventrolateral PFC, and especially in childhood (Kerr and Zelazo, 2004; Overman, 2004; Overman et al., 1996). Secondly, the comparison groups used, in the case of the children, attended the same schools for children with emotional and behavioural disorders, and in the case of the adults, were inmates in maximum security prisons. Whilst this might reduce other confounds such as school and home environment or socio-economic status, it also suggests that the comparisons do not form a 'healthy' group in the classic sense. An issue related to this is that both groups (especially in the child groups) demonstrated relatively low estimated IQ. It must be noted, however, that the measure used was an estimator of verbal IQ which has been negatively associated with antisocial behaviour. Importantly, however, there were no differences in IQ between the groups. A third issue, related to this, is that the comparisons may have presented with other disorders associated with antisocial behaviour such as anxiety or depression. Further studies should assess both target and comparison groups, in addition to the psychopathy measures, with a standard psychiatric interview. Related to this, ADHD was not measured in the adult sample. ADHD is purely a disorder of childhood in DSM-IV, however it is unlikely that ADHD simply disappears in adulthood (the issue of ADHD will be discussed further in section *6.5.1.*).

It would be valuable to use the current tasks in neuroimaging investigations of children with psychopathic tendencies and adult psychopaths. In particular,volumetric studies examining the size of amygdala and functional studies examining performance on these tasks would be useful. In relation to passive avoidance learning, given the increased dysfunction relating to

punishment-related learning it should be expected that an attenuated signal, *albeit within the same regions*, would be expected. On the basis of the results of experiment 3 and the predictions of the IES model, it might be predicted that a reduced signal would be observed in the amygdala and rostral ACC, especially during correct rejection trials (*i.e.* the signal might be stronger to hits). A complementary, attenuated 'punishment' signal might also be observed. In terms of reversal learning it would be expected that a psychopathic group would display a normal response in medial OFC, in relation to expectation violations, however the response in ventrolateral PFC may be expected to be attenuated during the commission of errors. Further, in a comparison of children with psychopathic tendencies and adult psychopaths during reversal learning a reduced signal, especially in the adults, might be expected. It would also be valuable to conduct pharmacological manipulations testing the performance on these tasks in order to attempt to simulate the psychopathic profile.

Finally, it would be interesting to test individuals with other disorders on the tasks used in this thesis. Individuals suffering with post traumatic stress disorder, depression, anxiety, intermittent explosive disorder, and paediatric bipolar disorder have all demonstrated increased propensities to react to frustration with aggression (*e.g.* Best et al, 2002; Blair, 2001a; Blair, 2004; Leibenluft et al., 2003). It is unlikely though that the underlying cause is the same in *all* of these disorders. Interestingly, individuals with depression, intermittent explosive disorder, and paediatric bipolar disorder have all also presented with impaired reversal learning ability (Best et al., 2002; Gorrindo et al., in press; Murphy et al., 2003). Notably, in the study by Best and colleagues (Best et al., 2002), performance on tasks attempting to asses DLPFC functioning were intact. This suggests that they may share similar aetiology, and thus present with similar neuroimaging profiles, as regards reversal learning. It must also be noted, however, that some disorders have been related to reversal learning impairment in the absence of increased tendencies towards reactive aggression, for example frontal variant fronto-temporal dementia (FvFTD) and Parkinson's disease, the latter being associated predominantly with striatal impairment (Swainson et al.,

2000). This cautions against pure 'frontal' accounts of 'executive functions' such as reversal learning (Elliott, 2003). It has been suggested that psychopathy, with regard to the emotional learning component, is almost the inverse of an anxiety disorder (Blair, 2004). In this sense it would be interesting to test individuals with anxiety disorders on the passive avoidance learning task.

6.5.1: The Impact of ADHD

ADHD has previously been associated with deficits in both passive avoidance and reversal learning (Iaboni et al., 1995; Itami and Uno, 2002; Milich et al., 1994). However the results of both experiments 1 and 4 indicated that the deficits on these tasks were associated with psychopathic tendencies even when level of ADHD was co-varied. Further ADHD was not found to be a significant covariate. In short, it appears that passive avoidance and reversal learning were preferentially impaired by the presence of psychopathic tendencies and not ADHD. In both experiments 1 and 4, however, the psychopathic tendencies group scored significantly higher on a measure of ADHD than did the comparison group. In fact, if the groups had been separated according to level of ADHD (disregarding psychopathic tendencies) group differences would have been observed – specifically, the high ADHD group would have demonstrated impaired passive avoidance and reversal learning compared with the low ADHD group.

What then accounts for the significant co-morbidity between the two disorders? As discussed in section *1.3.3.* there does not appear to be significant impairment in individuals with psychopathy on traditional measures of DLPFC functioning (LaPierre et al., 1995; Mitchell et al., 2002; Roussy and Toupin, 2000; Smith et al., 1992). Individuals with ADHD do, however, show difficulty with these tasks (Oosterlaan et al., 2005; Pennington and Ozonoff, 1996; Williams et al., 2000). Further whilst individuals with psychopathy do not demonstrate impairment on Stroop interference tasks (Newman et al., 1997; Peschardt et al., under revision; Smith et al., 1992), individuals with ADHD present with striking difficulty (Corbett and Stanczak, 1999; Leung and Connolly, 1996; Nigg et al., 2002; Pennington and Ozonoff, 1996; Reeve and Schandler, 2001). These

empirical results might be explained by differing neural bases purported to be involved in these two disorders. Whilst psychopathy has been associated with primary impairment in amygdala, ADHD has been associated with dysfunction of right-sided prefrontal-striatal systems (Casey et al., 1997; Castellanos et al., 1996; Giedd et al., 2001; Swanson et al., 1998). Indeed, Oosterlaan and colleagues (Oosterlaan et al., 2005) concluded that the presence of comborbid ADHD accounts for executive functioning impairments observed in children with disruptive behaviour disorders.

There have, however, been reports of common impairments between these two disorders. As suggested in section *1.3.3.* there is some evidence of impaired response control in individuals with psychopathy (LaPierre et al., 1995; Roussy and Toupin, 2000). Further, considerable data indicates that individuals with ADHD present with difficulty on response control paradigms (Berlin and Bohlin, 2002; Castellanos et al., 2000; Langley et al., 2004; Murphy, 2002; Oosterlaan et al., 1998; Oosterlaan et al., 2005; Oosterlaan and Sergeant, 1998a; Oosterlaan and Sergeant, 1998b; Pennington and Ozonoff, 1996; Rubia et al., 1998; Trommer et al., 1988; Wodushek and Neumann, 2003). Also, as noted in section *4.2.* there has been a report of impaired reversal learning in children with ADHD (Itami and Uno, 2002).

As demonstrated in experiment 6, neuroimaging investigations have critically associated reversal learning with ventrolateral PFC/lateral OFC (Cools et al., 2002; Kringelbach et al., 2003; O'Doherty et al., 2003a; O'Doherty et al., 2001; Remijnse et al., 2005). Similarly, neuroimaging studies have implicated this region in the successful performance of response control tasks (Casey et al., 2001; Garavan et al., 1999; Konishi et al., 1998; Konishi et al., 1999; Rubia et al., 2003). Thus, in explanation, it may be suggested that both children with psychopathic tendencies and children with ADHD share impairment within ventrolateral PFC, which leads to impaired reversal learning and response control. In short, it appears that ventrolateral PFC is involved in the regulation of on-line instrumental behaviour; particularly with respect to changing this behaviour following changes in contingency or task demands. It follows then that damage to

this region should be associated with difficulties in behavioural regulation, possibly manifesting as increased hyperactivity. Further, that that individuals with psychopathy presenting with this ventrolateral PFC dysfunction may also be at heightened risk of presenting with the impulsivity component of ADHD, and also that individuals with ADHD may, if this is associated with ventrolateral PFC dysfunction be at heightened risk for the display of reactive aggression. In accordance with these suggestions, as noted in section *1.2.2.*, it is the hyperactivity rather than the inattention component of ADHD that is associated with psychopathic tendencies (Colledge and Blair, 2001). In short, high co-morbidity of at least a hyperactive form of ADHD with psychopathic tendencies might be expected. Further, in keeping with the empirical data presented above, children with psychopathic tendencies would not be expected to present with the inattention component of ADHD.

The issue of ADHD is one that requires further study. It would be interesting to find 'pure' groups to test the predictions suggested in section *6.5.* For example, a pure 'inattentive' group would not be expected to perform poorly on either measure used in this thesis (or at least not for the same reasons as individuals with psychopathy), and would also not be expected to demonstrate abnormal responses within amygdala or ventrolateral PFC. In contrast, a 'hyperactive-impulsive' group would be expected to perform poorly on response control and reversal learning tasks, but to perform normally on the passive avoidance learning task. Further a pure 'callous-unemotional' group would be expected to perform abnormally only on the passive avoidance learning task, and to display defective amygdala responses. To find such groups may be difficult, or even impossible however, since impairment in one system over time might lead to impairment in others.

6.7: Conclusions

In conclusion, this thesis examined passive avoidance and probabilistic reversal learning in children with psychopathic tendencies and adult psychopaths. Both functions were impaired in these groups, however, it appears that the underlying neural networks required to perform each is largely different. It may be suggested that, in combination, the two underlying dysfunctions present in psychopathy, and investigated in this thesis, may form the bases of the two components within psychopathy. It has been suggested that the amygdala dysfunction is the primary, 'core', dysfunction within this disorder. Specifically, it appears that the emotion dysfunction, leading to the development of callous-unemotional traits and instrumental aggression (*i.e.* the presentation of factor 1 behaviours) may be exemplified by passive avoidance learning impairments. In addition, the dysfunction leading to the development of reactive aggression (*i.e.* behaviours relating to factor 2) may be exemplified by reversal learning impairments. The results in this thesis were largely compatible with the IES model of emotional learning.

References

Alm PO, Klinteberg B, Humble K, Leppert J, Sorensen S, Thorell LH, et al. Psychopathy, platelet MAO activity and criminality among former juvenile delinquents. Acta Psychiatrica Scandinavia 1996; 94: 105-11.

Amaral DG. The amygdaloid complex and the neurobiology of social behaviour. Society for Research in Child Development. Minneapolis, 2001.

Amaral DG, Price JL, Pitkanen A, Carmichael ST. Anatomical organization of the primate amygdaloid complex. In: Aggleton JP, editor. The Amygdala: Neurobiological aspects of emotion, memory, and mental dysfunction. New York: Wiley, 1992: 1-66.

Ambrogi Lorenzini C, Bucherelli C, Giachetti A, Mugani L, Tassoni G. Effects of nucleus basolateralis amygdalae neurotoxic lesions on aversive conditioning in the rat. Physiology and Behavior 1991; 49: 765-770.

Ambrogi Lorenzini CG, Baldi E, Bucherelli C, Sacchetti B, Tassoni G. Role of ventral hippocampus in acquisition, consolidation and retrieval of rat's passive avoidance response memory trace. Brain Research 1997; 768: 242-8.

Ambrogi Lorenzini CG, Baldi E, Bucherelli C, Sacchetti B, Tassoni G. Neural topography and chronology of memory consolidation: A review of functional inactivation findings. Neurobiology of Learning and Memory 1999; 71: 1-18.

Ambrogi Lorenzini CG, Baldi E, Bucherelli C, Tassoni G. Time-dependent deficits of rat's memory consolidation induced by tetrodotoxin injections into the caudate-putamen, nucleus accumbens, and globus pallidus. Neurobiology of Learning and Memory 1995; 63: 87-93.

Anderson SW, Bechara A, Damasio H, Tranel D, Damasio AR. Impairment of social and moral behaviour related to early damage in human prefrontal cortex. Nature Neuroscience 1999; 2: 1032-1037.

Angrilli, A., Mauri, A., Palomba, D., Flor, H., Birhaumer, N., Sartori, G. & di Paola, F. Startle reflex and emotion modulation impairment after a right amygdala lesion. Brain 1996; 119: 1991-2000.

Aniskiewicz AS. Autonomic components of vicarious conditioning and psychopathy. Journal of Clinical Psychology 1979; 35: 60-67.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders-IV. Washington, DC: APA, 1994.

Babinski LM, Hartsough CS, Lambert NM. Childhood conduct problems, hyperactivity-impulsivity, and inattention as predictors of adult criminal activity. Journal of Child Psychology and Psychiatry and allied disciplines 1999; 40: 347-355.

Baddeley AD, Della Sala S. Working memory and executive control. In A.C. Roberts, T.W. Robbins and L. Weiskrantz (Eds.) The Prefrontal Cortex: Executive and Cognitive Functions, Oxford: Oxford University Press, 1998.

Baird AA, Gruber SA, Fein DA, Maas LC, Steingard RJ, Renshaw PF, et al. Functional magnetic resonance imaging of facial affect recognition in children and adolescents. Journal of the American Academy of Child and Adolescent Psychiatry 1999; 38: 195-9.

Bandura A, Rosenthal TL. Vicarious classical conditioning as a function of arousal level. J Pers Soc Psychol 1966; 3: 54-62.

Barkley RA. Major life activity and health outcomes associated with attention-deficit/hyperactivity disorder. Journal of Clinical Psychiatry 2002; 63 Supplement 12: 10-5.

Barrash J, Tranel D, Anderson SW. Acquired personality disturbances associated with bilateral damage to the ventromedial prefrontal region. Developmental Neuropsychology 2000; 18: 355-81.

Barratt ES. Impulsiveness and aggression. In: Monahan J and Steadman H, editors. Violence and Mental Disorders: Developments in Risk Assessment. Chicago: University of Chicago Press, 1994: 61-79.

Barratt ES, Stanford MS, Dowdy L, Liebman MJ, Kent TA. Impulsive and premeditated aggression: a factor analysis of self- reported acts. Psychiatry Research 1999; 86: 163-73.

Barratt ES, Stanford MS, Felthous AR, Kent TA. The effects of phenytoin on impulsive and premeditated aggression: a controlled study. Journal of Clinical Psychopharmacology 1997a; 17: 341-9.

Barratt ES, Stanford MS, Kent TA, Felthous A. Neuropsychological and cognitive psychophysiological substrates of impulsive aggression. Biological Psychiatry 1997b; 41: 1045-61.

Barry CT, Frick PJ, DeShazo TM, McCoy MG, Ellis M, Loney BR. The importance of callous-unemotional traits for extending the concept of psychopathy to children. Journal of Abnormal Psychology 2000; 109: 335-40.

Baumrind D. Current patterns of parental authority. Developmental Psychology Monographs 1971; 4: 1-103.

Baumrind D. Rejoinder to Lewis's interpretation of parental firm control effects: Are authoritative families really harmonious? Psychological Bulletin 1983; 94: 132-142.

Baxter MG, Murray EA. The amygdala and reward. Nature Reviews. Neuroscience 2002; 3: 563-73.

Bechara A, Damasio AR, Damasio H, Anderson SW. Insensitivity to future consequences following damage to human prefrontal cortex. Cognition 1994; 50: 7-15.

Bechara, A., Tranel, D., Damasio, H., Adolphs, R., Rockland, C. & Damasio, A. R. Double dissociation of conditioning and declarative knowledge relative to the amygdala and hippocampus in humans. Science 1995; 269: 1115-1118.

Bechara A, Damasio H, Damasio AR. Emotion, decision making and the orbitofrontal cortex. Cerebral Cortex 2000a; 10: 295-307.

Bechara A, Damasio H, Damasio AR, Lee GP. Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. Journal of Neuroscience 1999; 19: 5473-5481.

Bechara A, Damasio H, Tranel D, Damasio AR. Deciding advantageously before knowing the advantageous strategy. Science 1997; 275: 1293-1295.

Bechara A, Dolan S, Denburg N, Hindes A, Anderson SW, Nathan PE. Decision-making deficits, linked to a dysfunctional ventromedial prefrontal cortex, revealed in alcohol and stimulant abusers. Neuropsychologia 2001; 39: 376-89.

Bechara A, Tranel D, Damasio H. Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. Brain 2000b; 123 (Pt 11): 2189-202.

Berkowitz L. Aggression: Its causes, consequences, and control. Philadelphia: Temple University Press, 1993.

Berlin HA, Rolls ET, Kischka U. Impulsivity, time perception, emotion and reinforcement sensitivity in patients with orbitofrontal cortex lesions. Brain 2004; 127: 1108-26.

Berlin L, Bohlin G. Response inhibition, hyperactivity, and conduct problems among preschool children. Journal of Clinical Child and Adolescent Psychology 2002; 31: 242-51.

Bermudez-Rattoni F, Introini-Collison I, Coleman-Mesches K, McGaugh JL. Insular cortex and amygdala lesions induced after aversive training impair retention: effects of degree of training. Neurobiology of Learning and Memory 1997; 67: 57-63.

Bermudez-Rattoni F, Introini-Collison IB, McGaugh JL. Reversible inactivation of the insular cortex by tetrodotoxin produces retrograde and anterograde amnesia for inhibitory avoidance and spatial learning. Proceedings of the National Academy of Sciences of the United States of America 1991; 88: 5379-82.

Bermudez-Rattoni F, McGaugh JL. Insular cortex and amygdala lesions differentially affect acquisition on inhibitory avoidance and conditioned taste aversion. Brain Research 1991; 549: 165-70.

Berthoz S, Armony J, Blair RJR, Dolan R. Neural correlates of violation of social norms and embarrassment. Brain 2002; 125: 1696-1708.

Best M, Williams JM, Coccaro EF. Evidence for a dysfunctional prefrontal circuit in patients with an impulsive aggressive disorder. Proceedings of the National Academy of Sciences of the United States of America 2002; 99: 8448-53.

Biederman J, Newcorn J, Sprich S. Comorbidity of attention deficit hyperactivity disorder with conduct, depressive, anxiety, and other disorders. American Journal of Psychiatry 1991; 148: 564-577.

Blackburn R. On moral judgements and personality disorders: The myth of psychopathic personality revisited. British Journal of Psychiatry 1988; 153: 505-512.

Blair, Colledge, Mitchell. Somatic markers and response reversal: is there orbitofrontal cortex dysfunction in boys with psychopathic tendencies? Journal of Abnormal Child Psychology 2001a; 29: 499-511.

Blair HT, Schafe GE, Bauer EP, Rodrigues SM, LeDoux JE. Synaptic plasticity in the lateral amygdala: a cellular hypothesis of fear conditioning. Learning and Memory 2001b; 8: 229-42.

Blair RJ. A cognitive developmental approach to mortality: investigating the psychopath. Cognition 1995; 57: 1-29.

Blair RJ. Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy. Journal of Neurology Neurosurgery and Psychiatry 2001; 71: 727-31.

Blair RJ. Neurobiological basis of psychopathy. British Journal of Psychiatry 2003a; 182: 5-7.

Blair RJ. The roles of orbital frontal cortex in the modulation of antisocial behavior. Brain and Cognition 2004; 55: 198-208.

Blair RJ, Budhani S, Colledge E, Scott S. Deafness to fear in boys with psychopathic tendencies. Journal of Child Psychology and Psychiatry 2005; 46: 327-36.

Blair RJ, Cipolotti L. Impaired social response reversal. A case of 'acquired sociopathy'. Brain 2000; 123 (Pt 6): 1122-41.

Blair RJ, Colledge E, Murray L, Mitchell DG. A selective impairment in the processing of sad and fearful expressions in children with psychopathic tendencies. Journal of Abnormal Child Psychology 2001c; 29: 491-8.

Blair RJ, Jones L, Clark F, Smith M. The psychopathic individual: a lack of responsiveness to distress cues? Psychophysiology 1997; 34: 192-8.

Blair RJ, Mitchell DG, Richell RA, Kelly S, Leonard A, Newman C, et al. Turning a deaf ear to fear: impaired recognition of vocal affect in psychopathic individuals. Journal of Abnormal Psychology 2002; 111: 682-6.

Blair RJR. Moral reasoning in the child with psychopathic tendencies. Personality and Individual Differences 1997; 22: 731-739.

Blair RJR. Responsiveness to distress cues in the child with psychopathic tendencies. Personality and Individual Differences 1999; 27: 135-145.

Blair RJR. A neurocognitive model of the psychopathic individual. In: Ron MA and Robbins TW, editors. Disorders of Brain and Mind 2. Cambridge: Cambridge University Press, 2003b: 400-420.

Blair RJR, Coles M. Expression recognition and Behavioural problems in early adolescence. Cognitive Development 2000; 15: 421-434.

Blair RJR, Jones L, Clark F, Smith M. Is the psychopath "morally insane"? Personality and Individual Differences 1995a; 19: 741-752.

Blair RJR, Mitchell DGV, Leonard A, Budhani S, Peschardt KS, Newman C. Passive avoidance learning in psychopathic individuals: Modulation by reward but not by punishment. Personality and Individual Differences. 2004; 37: 1179-1192.

Blair RJR, Monson J, Frederickson N. Moral reasoning and conduct problems in children with emotional and behavioural difficulties. Personality and Individual Differences 2001d; 31: 799-811.

Blair RJR, Morris JS, Frith CD, Perrett DI, Dolan R. Dissociable neural responses to facial expressions of sadness and anger. Brain 1999; 122: 883-893.

Blair RJR, Morton J. Putting cognition into sociopathy. Brain and Behavioral Science 1995; 18: 548.

Blair RJR, Sellars C, Strickland I, Clark F, Williams AO, Smith M, et al. Emotion attributions in the psychopath. Personality and Individual Differences 1995b; 19: 431-437.

Blanchard RJ, Blanchard DC, Takahashi LK. Attack and defensive behaviour in the albino rat. Animal Behavior 1977; 25: 197-224.

Blumstein A, Cohen J. Characterizing criminal careers. Science 1987; 237: 985-991.

Blundell P, Hall G, Killcross S. Preserved sensitivity to outcome value after lesions of the basolateral amygdala. Journal of Neuroscience 2003; 23: 7702-9.

Bohn I, Giertler C, Hauber W. NMDA receptors in the rat orbital prefrontal cortex are involved in guidance of instrumental behaviour under reversal conditions. Cerebral Cortex 2003a; 13: 968-76.

Bohn I, Giertler C, Hauber W. Orbital prefrontal cortex and guidance of instrumental behaviour in rats under reversal conditions. Behavioural and Brain Research 2003b; 143: 49-56.

Botvinick M, Nystrom LE, Fissell K, Carter CS, Cohen JD. Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature 1999; 402: 179-81.

Botvinick MM, Cohen JD, Carter CS. Conflict monitoring and anterior cingulate cortex: an update. Trends in Cognitive Sciences 2004; 8: 539-46.

Breiter HC, Aharon I, Kahneman D, Dale A, Shizgal P. Functional imaging of neural responses to expectancy and experience of monetary gains and losses. Neuron 2001; 30: 619-39.

Breiter HC, Etcoff NL, Whalen PJ, Kennedy WA, Rauch SL, Buckner RL, et al. Response and habituation of the human amygdala during visual processing of facial expression. Neuron 1996; 17: 875-887.

Brody GH, Shaffer DR. Contributions of parents and peers to children's moral socialisation. Developmental Review 1982; 2: 31-75.

Burgess PW, Alderman N, Evans J, Emslie H, Wilson BA. The ecological validity of tests of executive function. Journal of the International Neuropsychological Society 1998; 4, 547-558.

Burgess PW, Wood RL. Neuropsychology of behaviour disorders following brain injury. In: Wood RL, editor. Neurobehavioural Sequelae of Traumatic Brain Injury. London: Taylor & Francis, 1990: 110-133.

Busemeyer JR, Stout JC. A contribution of cognitive decision models to clinical assessment: decomposing performance on the Bechara gambling task. Psychol Assess 2002; 14: 253-62.

Bush G, Frazier JA, Rauch SL, Seidman LJ, Whalen PJ, Jenike MA, et al. Anterior cingulate cortex dysfunction in attention-deficit/hyperactivity disorder revealed by fMRI and the Counting Stroop. Biological Psychiatry 1999; 45: 1542-52.

Bush G, Luu P, Posner MI. Cognitive and emotional influences in anterior cingulate cortex. Trends in Cognitive Sciences 2000; 4: 215-222.

Cahill L, McGaugh JL. Amygdaloid complex lesions differentially affect retention of tasks using appetitive and aversive reinforcement. Behavioural Neuroscience 1990; 104: 532-43.

Cale EM, Lilienfeld SO. Sex differences in psychopathy and antisocial personality disorder. A review and integration. Clinical Psychology Review 2002; 22: 1179-207.

Carter CS, Braver TS, Barch DM, Botvinick MM, Noll D, Cohen JD. Anterior cingulate cortex, error detection, and the online monitoring of performance. Science 1998; 280: 747-9.

Casey BJ, Castellanos FX, Giedd JN, Marsh WL, Hamburger SD, Schubert AB, et al. Implication of right frontostriatal circuitry in response inhibition and

attention-deficit/hyperactivity disorder. Joural of the American Academy of Child and Adolescent Psychiatry 1997; 36: 374-83.

Casey BJ, Forman SD, Franzen P, Berkowitz A, Braver TS, Nystrom LE, et al. Sensitivity of prefrontal cortex to changes in target probability: a functional MRI study. Human Brain Mapping 2001; 13: 26-33.

Casey BJ, Thomas KM, Davidson MC, Kunz K, Franzen PL. Dissociating striatal and hippocampal function developmentally with a stimulus-response compatibility task. Journal of Neuroscience 2002; 22: 8647-52.

Casey BJ, Thomas KM, Welsh TF, Badgaiyan RD, Eccard CH, Jennings JR, et al. Dissociation of response conflict, attentional selection, and expectancy with functional magnetic resonance imaging. Proceedings of the National Academy of Sciences of the United States of America 2000; 97: 8728-33.

Castellanos FX, Giedd JN, Marsh WL, Hamburger SD, Vaituzis AC, Dickstein DP, et al. Quantitative brain magnetic resonance imaging in attention-deficit hyperactivity disorder. Archives of General Psychiatry 1996; 53: 607-616.

Castellanos FX, Marvasti FF, Ducharme JL, Walter JM, Israel ME, Krain A, et al. Executive function oculomotor tasks in girls with ADHD. Journal of the American Academy of Child and Adolescent Psychiatry 2000; 39: 644-50.

Christian RE, Frick PJ, Hill NL, Tyler L, Frazer DR. Psychopathy and conduct problems in children: II. Implications for subtyping children with conduct problems. Journal of the American Academy of Child and Adolescent Psychiatry 1997; 36: 233-41.

Church RM. Emotional reactions of rats to the pain of others. Journal of Comparative & Physiological Psychology 1959; 52: 132-134.

Clarke HF, Dalley JW, Crofts HS, Robbins TW, Roberts AC. Cognitive inflexibility after prefrontal serotonin depletion. Science 2004; 304: 878-80.

Clarke HF, Walker SC, Crofts HS, Dalley JW, Robbins TW, Roberts AC. Prefrontal serotonin depletion affects reversal learning but not attentional set shifting. Journal of Neuroscience 2005; 25: 532-8.

Cleckley HM. The mask of sanity. St Louis, MO: Mosby, 1941.

Cleckley HM. The Mask of Sanity. 5th Edition. St Louis, MO: Mosby, 1976.

Cohen MS. Parametric analysis of fMRI data using linear systems methods. Neuroimage 1997; 6: 93-103.

Colledge E, Blair RJR. Relationship between Attention-Deficit-Hyperactivity Disorder and Psychopathic Tendencies in Children. Personality and Individual Differences 2001; 30: 1175-1187.

Cooke DJ, Michie C. Refining the construct of psychopathy: Towards a hierarchical model. Psychological Assessment 2001; 13: 171-188.

Cools R, Barker RA, Sahakian BJ, Robbins TW. Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. Cerebral Cortex 2001; 11: 1136-43.

Cools R, Clark L, Owen AM, Robbins TW. Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. Journal of Neuroscience 2002; 22: 4563-7.

Corbett B, Stanczak DE. Neuropsychological performance of adults evidencing attention-deficit hyperactivity disorder. Archives of Clinical Neuropsychology 1999; 14: 373-87.

Cornell DG, Warren J, Hawk G, Stafford E, Oram G, Pine D. Psychopathy in instrumental and reactive violent offenders. Journal of Consulting and Clinical Psychology 1996; 64: 783-790.

Cox RW. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research 1996; 29: 162-73.

Cox SM, Andrade A, Johnsrude IS. Learning to like: a role for human orbitofrontal cortex in conditioned reward. Journal of Neuroscience 2005; 25: 2733-40.

Crick NR, Dodge KA. Social information-processing mechanisms on reactive and proactive aggression. Child Development 1996; 67: 993-1002.

Dale E, Reichart D. Bibliography of Vocabulary Studies. Columbus, OH: Ohio State University, Bureau of Educational Research, 1957.

Damasio AR. Descartes' error: Emotion, rationality and the human brain. New York: Putnam (Grosset Books), 1994.

Damasio AR, Tranel D, Damasio H. Individuals with sociopathic behaviour caused by frontal damage fail to respond autonomically to social stimuli. Behavioural Brain Research 1990; 41: 81-94.

Damasio AR, Tranel D, Damasio HC. Somatic markers and the guidance of behavior: Theory and preliminary testing. In: Levin HS, Eisenberg HM and Benton AL, editors. Frontal Lobe Function and Dysfunction. New York: Oxford University Press, 1991: 217-229.

Deacon RM, Penny C, Rawlins JN. Effects of medial prefrontal cortex cytotoxic lesions in mice. Behavioural Brain Research 2003; 139: 139-55.

Delgado MR, Locke HM, Stenger VA, Fiez JA. Dorsal striatum responses to reward and punishment: effects of valence and magnitude manipulations. Cognitive, Affective and Behavioral Neuroscience 2003; 3: 27-38.

Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA. Tracking the hemodynamic responses to reward and punishment in the striatum. Journal of Neurophysiology 2000; 84: 3072-7.

Desimone R, Duncan J. Neural mechanisms of selective visual attention. Annual Review of Neuroscience 1995; 18: 193-222.

Dias R, Robbins TW, Roberts AC. Dissociation in prefrontal cortex of affective and attentional shifts. Nature 1996; 380: 69-72.

Dias R, Robbins TW, Roberts AC. Dissociable forms of inhibitory control within prefrontal cortex with an analog of the Wisconsin Card Sort Test: restriction to novel situations and independence from "on-line" processing. Journal of Neuroscience 1997; 17: 9285-97.

Dickinson A. Contemporary Animal Learning Theory. Cambridge: Cambridge University Press, 1980.

Divac I, Rosvold HE, Szwarcbart MK. Behavioral effects of selective ablation of the caudate nucleus. Journal of Comparative Physiology and Psychology 1967; 63: 184-90.

Dodd T, Nicholas S, Povey D, Walker A. Crime in England and Wales 2003/2004. London, 2004.

Dodge KA, Coie JD. Social-information-processing factors in reactive and proactive aggression in children's peer groups. Journal of Personality and Social Psychology 1987; 53: 1146-58.

Dolan M, Park I. The neuropsychology of antisocial personality disorder. Psychological Medicine 2002; 32: 417-27.

Dolan MC, Anderson IM. The relationship between serotonergic function and the Psychopathy Checklist: Screening Version. Journal of Psychopharmacology 2003; 17: 216-22.

Dougherty DD, Shin LM, Alpert NM, Pitman RK, Orr SP, Lasko M, et al. Anger in healthy men: a PET study using script-driven imagery. Biological Psychiatry 1999; 46: 466-72.

Downes JJ, Roberts AC, Sahakian BJ, Evenden JL, Morris RG, Robbins TW. Impaired extra-dimensional shift performance in medicated and unmedicated Parkinson's disease: evidence for a specific attentional dysfunction. Neuropsychologia 1989; 27: 1329-43.

Drevets WC, Lowry T, Gautier C, Perrett DI, Kupfer DJ. Amygdalar blood flow responses to facially expressed sadness. Biological Psychiatry 2000; 47: 160S.

Dunn LM, Wheklan C, Pintillie D. British Picture Vocabulary Scale. Windsor, UK: NFER-Nelson, 1982.

DuPaul GJ. Parent and teacher ratings of ADHD symptoms: Psychometric properties in a community-based sample. Journal of Clinical Child Psychology 1991; 20: 245-253.

DuPaul GJ, Power TJ, Anastopoulos AD, Reid R. ADHD Rating Scale--IV: Checklists, norms, and clinical interpretation. New York, NY, US: The Guilford Press, 1998.

Eisenberg N, Fabes RA, Guthrie IK, Murphy BC, Maszk P, Holmgren R, et al. The relations of regulation and emotionality to problem behaviour in

elementary school children. Development and Psychopathology 1996; 8: 141-162.

Elliot FA. Neurological aspects of antisocial behavior. In: Reid WH, editor. The psychopath. New York: Bruner/ Mazel, 1978.

Elliott R. Executive functions and their disorders. British Medical Bulletin 2003; 65: 49-59.

Elliott R, Dolan RJ, Frith CD. Dissociable functions in the medial and lateral orbitofrontal cortex: evidence from human neuroimaging studies. Cerebral Cortex 2000a; 10: 308-17.

Elliott R, Friston KJ, Dolan RJ. Dissociable neural responses in human reward systems. Journal of Neuroscience 2000b; 20: 6159-65.

Elliott R, Newman JL, Longe OA, William Deakin JF. Instrumental responding for rewards is associated with enhanced neuronal response in subcortical reward systems. Neuroimage 2004; 21: 984-90.

Everitt BJ, Cardinal RN, Hall J, Parkinson JA, Robbins TW. Differential involvement of amygdala subsystems in appetitive conditioning and drug addiction. In: Aggleton JP, editor. The Amygdala: A functional analysis. Oxford: Oxford University Press, 2000: 289-310.

Everitt BJ, Cardinal RN, Parkinson JA, Robbins TW. Appetitive behavior: impact of amygdala-dependent mechanisms of emotional learning. Annals of the New York Academy of Sciences 2003; 985: 233-50.

Eysenck HJ. Crime and personality. London: Routledge & Kegan Paul, 1964.

Eysenck HJ, Gudjonsson GH. The causes and cures of criminality. London: Plenum Press, 1989.

Farrington DP. Age and crime. In: Tonry M and Morris N, editors. Crime and Justice: An Annual Review of Research. Vol 7. Chicago: University of Chicago Press, 1986: 189-250.

Farrington DP. Implications of criminal career research for the prevention of offending. Journal of Adolescence 1990; 13: 93-113.

Farrington DP, West DJ. Criminal, penal and life histories of chronic offenders: risk and protective factors and early identification. Criminal Behaviour and Mental Health 1993; 3: 492-523.

Fazel S, Danesh J. Serious mental disorder in 23000 prisoners: a systematic review of 62 surveys. Lancet 2002; 359: 545-50.

Fellows LK, Farah MJ. Ventromedial frontal cortex mediates affective shifting in humans: evidence from a reversal learning paradigm. Brain 2003; 126: 1830-7.

Fellows LK, Farah MJ. Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. Cerebral Cortex 2005; 15: 58-63.

Feshbach ND. Parental empathy and child adjustment/ maladjustment. In: Eisenberg N and Strayer J, editors. Empathy and its development. New York: Cambridge University Press, 1987.

Fine, C. Expectation violations and emotional learning. 2000. London, University College London, University of London.

Fine, C. & Blair, R. J. R. Mini review: The cognitive and emotional effects of amygdala damage. Neurocase 2000: 435-450.

Fine C, Richell RA, Mitchell DGV, Newman C, Lumsden J, Blair RJR. Instrumental learning and response reversal: The involvement of the amygdala and orbital frontal cortex and implications for psychopathy. submitted.

Fisher L, Blair RJR. Cognitive impairment and its relationship to psychopathic tendencies in children with emotional and behavioural difficulties. Journal of Abnormal Child Psychology 1998; 26: 511-519.

Flor H, Birbaumer N, Hermann C, Ziegler S, Patrick CJ. Aversive Pavlovian conditioning in psychopaths: Peripheral and central correlates. Psychophysiology 2002; 39: 505-518.

Forth AE, Kosson DS, Hare RD. The Psychopathy Checklist: Youth Version. Toronto, Ontario, Canada: Multi-Health Systems, in press.

Fowles DC. Psychophysiology and Psychopathy: A motivational approach. Psychophysiology 1988; 25: 373-391.

Frick PJ. Callous-unemotional traits and conduct problems: a two-factor model of psychopathy in children. Issues in Criminological and Legal Psychology 1995; 24: 47-51.

Frick PJ. The problems of internal validation without a theoretical context: the different conceptual underpinnings of psychopathy and the disruptive behavior disorder criteria. Psychological Assessment 2000; 12: 451-6.

Frick PJ, Bodin SD, Barry CT. Psychopathic traits and conduct problems in community and clinic-referred samples of children: further development of the psychopathy screening device. Psychological Assessment 2000; 12: 382-93.

Frick PJ, Ellis M. Callous-unemotional traits and subtypes of conduct disorder. Clinical, Child and Family Psychology Review 1999; 2: 149-68.

Frick PJ, Hare RD. Antisocial Process Screening Device. Toronto: Multi-Health Systems, 2001.

Frick PJ, O'Brien BS, Wootton JM, McBurnett K. Psychopathy and Conduct Problems in Children. Journal of Abnormal Psychology 1994a; 103: 700-707.

Fuster JM. The Prefrontal Cortex. New York: Raven, 1980.

Gallagher M, McMahan RW, Schoenbaum G. Orbitofrontal cortex and representation of incentive value in associative learning. Journal of Neuroscience 1999; 19: 6610-6614.

Garavan H, Ross TJ, Stein EA. Right hemispheric dominance of inhibitory control: an event-related functional MRI study. Proceedings of the National Academy of Sciences of the United States of America 1999; 96(14): 8301-8306.

Garavan H, Ross TJ, Murphy K, Roche RA, Stein EA. Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction. Neuroimage 2002; 17: 1820-9.

Giedd JN, Blumenthal J, Molloy E, Castellanos FX. Brain imaging of attention deficit/hyperactivity disorder. Annals of the New York Academy of the Sciences 2001; 931: 33-49.

Gorenstein EE. Frontal lobe functions in psychopaths. Journal of Abnormal Psychology 1982; 91: 368-379.

Gorenstein EE, Newman JP. Disinhibitory Psychopathology: A new perspective and a model for research. Psychological Review 1980; 37: 301-315.

Gorrindo T, Blair RJR, Budhani S, Dickstein D, Pine D, Leibenluft E. Probabilistic response reversal deficts in pediatric bipolar disorder. American Journal of Psychiatry. in press.

Goveas JS, Csernansky JG, Coccaro EF. Platelet serotonin content correlates inversely with life history of aggression in personality-disordered subjects. Psychiatry Research 2004; 126: 23-32.

Goyer PF, Andreason PJ, Semple WE, Clayton AH, King AC, Compton-Toth BA, et al. Positron-emission tomography and personality disorders. Neuropsychopharmacology 1994; 10: 21-8.

Grafman J, Schwab K, Warden D, Pridgen BS, Brown HR. Frontal lobe injuries, violence, and aggression: A report of the Vietnam head injury study. Neurology 1996; 46: 1231-1238.

Grann M, Langstrom N, Tengstrom A, Kullgren G. Psychopathy (PCL-R) predicts violent recidivism among criminal offenders with personality disorders in Sweden. Law and Human Behavior 1999; 23: 205-17.

Gray JA. The structure of the emotions and the limbic system. Ciba Foundation Symposium 1972; 8: 87-120.

Gray JA. The psychology of fear and stress. Cambridge: University of Cambridge Press, 1987.

Gray JA, McNaughton N. The neuropsychology of anxiety: reprise. Nebraska Symposium on Motivation 1996; 43: 61-134.

Gregg TR, Siegel A. Brain structures and neurotransmitters regulating aggression in cats: implications for human aggression. Progress in Neuropsychopharmacology and Biological Psychiatry 2001; 25: 91-140.

Gutierrez H, Hernandez-Echeagaray E, Ramirez-Amaya V, Bermudez-Rattoni F. Blockade of N-methyl-D-aspartate receptors in the insular cortex disrupts taste aversion and spatial memory formation. Neuroscience 1999; 89: 751-8.

Hare RD. Temporal gradient of fear arousal in psychopaths. Journal of Abnormal Psychology 1965; 70: 442-445.

Hare RD. Psychopathy and physiological activity during anticipation of an aversive stimulus in a distraction paradigm. Psychophysiology 1982; 19: 266-271.

Hare RD. The Hare Psychopathy Checklist-Revised. Toronto, Ontario: Multi-Health Systems, 1991.

Hare RD. The Hare Psychopathy Checklist-Revised: 2nd Edition. Toronto, Ontario: Multi-Health Systems, 2003.

Hare RD, Clark D, Grann M, Thornton D. Psychopathy and the predictive validity of the PCL-R: an international perspective. Behavioral Sciences and the Law 2000; 18: 623-45.

Hare RD, Frazelle J, Cox DN. Psychopathy and physiological responses to threat of an aversive stimulus. Psychophysiology 1978; 15: 165-172.

Hare RD, McPherson LM. Psychopathy and perceptual asymmetry during verbal dichotic listening. J Abnormal Psychology 1984; 93: 141-9.

Hare RD, Quinn MJ. Psychopathy and autonomic conditioning. Journal of Abnormal Psychology 1971; 77: 223-235.

Harmer CJ, Perrett DI, Cowen PJ, Goodwin GM. Administration of the beta-adrenoceptor blocker propranolol impairs the processing of facial expressions of sadness. Psychopharmacology (Berl) 2001; 154: 383-9.

Harpur TJ, Hakstian AR, Hare RD. The factor structure of the Psychopathy Checklist. Journal of Consulting and Clinical Psychology 1988; 56: 741-747.

Harpur TJ, Hare RD, Hakstian AR. Two-factor conceptualization of psychopathy: Construct validity and assessment implications. Psychological

Assessment: A Journal of Consulting and Clinical Psychology 1989; 1: 6-17.

Hart S, Kropp PR, Hare RD. Performance of male psychopaths following conditional release from prison. Journal of Consulting and Clinical Psychology 1988; 56: 227-232.

Hart SD, Hare RD. Psychopathy and antisocial personality disorder. Current Opinion in Psychiatry 1996; 9: 129-132.

Hart SD, Hare RD. Psychopathy: Assessment and association with criminal conduct. In: Stoff DM and Breiling J, editors. Handbook of Antisocial Behaviour. New York, NY, US: John Wiley & Sons, Inc, 1997.

Hartung CM, Milich R, Lynam DR, Martin CA. Understanding the relations among gender, disinhibition, and disruptive behavior in adolescents. Journal of Abnormal Psychology 2002; 111: 659-64.

Hawes DJ, Dadds MR. The Treatment of Conduct Problems in Children with Callous-Unemotional Traits. Journal of Consulting and Clinical Psychology in press.

Hebb DO. The organization of behavior. New York: John Wiley & Sons, 1949.

Hecaen H, Albert ML. Human Neuropsychology. New York: Wiley, 1978.

Hemphill JF, Hare RD, Wong S. Psychopathy and recidivism: A review. Legal and Criminological Psychology 1998; 3: 139-170.

Hemphill JF, Hart SD, Hare RD. Psychopathy and substance use. Journal of Personality Disorders 1994; 8: 169-180.

Hinshaw SP. On the distinction between attentional deficits/hyperactivity and conduct problems/aggression in child psychopathology. Psychological Bulletin 1987; 101: 443-463.

Hoffman ML. Empathy, its limitations, and its role in a comprehensive moral theory. In: Gewirtz J and Kurtines W, editors. Morality, Moral Development, and Moral Behavior. New York: Wiley, 1984: 283-302.

Hoffman ML. Discipline and internalisation. Developmental Psychology 1994; 30: 26-28.

Hoffman ML, Saltzstein HD. Parent discipline and the child's moral development. Journal of Personality and Social Psychology 1967; 5: 45-57.

Hornak J, O'Doherty J, Bramham J, Rolls ET, Morris RG, Bullock PR, et al. Reward-related reversal learning after surgical excisions in orbito-frontal or dorsolateral prefrontal cortex in humans. Journal of Cognitive Neuroscience 2004; 16: 463-78.

House TH, Milligan WL. Autonomic responses to modeled distress in prison psychopaths. Journal of Personality and Social Psychology 1976; 34: 556-560.

Iaboni F, Douglas VI, Baker AG. Effects of reward and response costs on inhibition in ADHD children. Journal of Abnormal Psychology 1995; 104: 232-40.

Itami S, Uno H. Orbitofrontal cortex dysfunction in attention-deficit hyperactivity disorder revealed by reversal and extinction tasks. Neuroreport 2002; 13: 2453-7.

Iversen SD, Mishkin M. Perseverative interference in monkeys following selective lesions of the inferior prefrontal convexity. Experimental Brain Research 1970; 11: 376-86.

Izquierdo A, Suda RK, Murray EA. Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. Journal of Neuroscience 2004; 24: 7540-8.

Johansson P, Kerr M, Andershed H. Linking adult psychopathy with childhood hyperactivity-impulsivity-attention problems and conduct problems through retrospective self-reports. Journal of Personality Disorders 2005; 19: 94-101.

Kahneman D, Tversky A. Prospect theory: an analysis of decision under risk. Econometrica 1979; 47: 263-292.

Kandel E, Freed D. Frontal lobe dysfunction and antisocial behavior: a review. Journal of Clinical Psychology 1989; 45: 404-413.

Kerns JG, Cohen JD, MacDonald AW, 3rd, Cho RY, Stenger VA, Carter CS. Anterior cingulate conflict monitoring and adjustments in control. Science 2004; 303: 1023-6.

Kerr A, Zelazo PD. Development of "hot" executive function: the children's gambling task. Brain and Cognition 2004; 55: 148-57.

Kiehl KA, Hare RD, McDonald JJ, Brink J. Semantic and affective processing in psychopaths: An event-related potential (ERP) study. Psychophysiology 1999; 36: 765–774.

Killcross S, Robbins TW, Everitt BJ. Different types of fear-conditioned behaviour mediated by separate nuclei within amygdala. Nature 1997; 388: 377-80.

Knutson B, Adams CM, Fong GW, Hommer D. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. Journal of Neuroscience 2001; 21: RC159.

Konishi S, Nakajima K, Uchida I, Sekihara K, Miyashita Y. No-go dominant brain activity in human inferior prefrontal cortex revealed by functional magnetic resonance imaging. European Journal of Neuroscience 1998; 10: 1209-13.

Konishi S, Nakajima K, Uchida I, Kikyo H, Kameyama M, Miyashita Y. Brain 1999; 122(5): 981-991.

Kosson DS, Cyterski TD, Steuerwald BL, Neumann CS, Walker-Matthews S. The reliability and validity of the psychopathy checklist: youth version (PCL:YV) in nonincarcerated adolescent males. Psychological Assessment 2002; 14: 97-109.

Kosson DS, Smith SS, Newman JP. Evaluating the construct validity of psychopathy in black and white male inmates: three preliminary studies. Journal of Abnormal Psychology 1990; 99: 250-9.

Kringelbach ML, O'Doherty J, Rolls ET, Andrews C. Activation of the human orbitofrontal cortex to a liquid food stimulus is correlated with its subjective pleasantness. Cerebral Cortex 2003; 13: 1064-71.

Kringelbach ML, Rolls ET. Neural correlates of rapid reversal learning in a simple model of human social interaction. Neuroimage 2003; 20: 1371-83.

Kringelbach ML, Rolls ET. The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology. Progressive Neurobiology 2004; 72: 341-72.

LaBar, K. S., LeDoux, J. E., Spencer, D. D. & Phelps, E. A. Impaired fear conditioning following unilateral temporal lobectomy in humans. 1995 15: 6846-6855.

Lahey BB, Goodman SH, Waldman ID, Bird H, Canino G, Jensen P, et al. Relation of age of onset to the type and severity of child and adolescent conduct problems. Journal of Abnormal Child Psychology 1999; 27: 247-60.

Lang PJ, Bradley MM, Cuthbert BN. Emotion, attention, and the startle reflex. Psychological Review 1990; 97: 377-398.

Langley K, Marshall L, Van Den Bree M, Thomas H, Owen M, O'Donovan M, et al. Association of the Dopamine D(4) Receptor Gene 7-Repeat Allele With Neuropsychological Test Performance of Children With ADHD. American Journal of Psychiatry 2004; 161: 133-138.

LaPierre D, Braun CMJ, Hodgins S. Ventral frontal deficits in psychopathy: Neuropsychological test findings. Neuropsychologia 1995; 33: 139-151.

Lavie N. Perceptual load as a necessary condition for selective attention. Journal of Experimental Psychology: Human Perception and Performance 1995; 21: 451-68.

Lawrence AD, Sahakian BJ, Rogers RD, Hodge JR, Robbins TW. Discrimination, reversal, and shift learning in Huntington's disease: mechanisms of impaired response selection. Neuropsychologia 1999; 37: 1359-74.

LeDoux JE. The amygdala and emotion: a view through fear. In: Aggleton JP, editor. The Amygdala: A functional analysis. Oxford: Oxford University Press, 2000: 289-310.

Leibenluft E, Blair RJ, Charney DS, Pine DS. Irritability in pediatric mania and other childhood psychopathology. Annals of the New York Academy of Sciences 2003; 1008: 201-18.

Leung PWL, Connolly KJ. Distractibility in Hyperactive and Conduct Disordered Children. Journal of Child Psychology and Psychiatry 1996; 37: 305-312.

Levenston GK, Patrick CJ, Bradley MM, Lang PJ. The psychopath as observer: Emotion and attention in picture processing. Journal of Abnormal Psychology 2000; 109: 373-386.

Liddle PF, Kiehl KA, Smith AM. Event-related fMRI study of response inhibition. Human Brain Mapping 2001; 12: 100-9.

Linnoila M, Virkkunen M, Scheinin M, Nuutila A, Rimon R, Goodwin FK. Low cerebrospinal fluid 5-hydroxy indoleacetic acid concentration differentiates impulsive from nonimpulsive violent behavior. Life Sciences 1983; 33: 2609-2614.

Loeber R, Farrington DP. Young children who commit crime: epidemiology, developmental origins, risk factors, early interventions, and policy implications. Development and Psychopathology 2000; 12: 737-62.

Lorenz AR, Newman JP. Deficient response modulation and emotion processing in low-anxious caucasian psychopathic offenders: Results from a lexical decision task. Emotion 2002; 2: 91-104.

Luria AR. Human Brain and Psychological Processes. New York: Harper and Row, 1966.

Lykken DT. A study of anxiety in the sociopathic personality. Journal of Abnormal and Social Psychology 1957; 55: 6-10.

Lykken DT. The Antisocial Personalities. Hillsdale, New Jersey: Erlbaum, 1995.

Lynam DR. Early identification of chronic offenders: who is the fledgling psychopath? Psychological Bulletin 1996; 120.

MacDonald AW, 3rd, Cohen JD, Stenger VA, Carter CS. Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. Science 2000; 288: 1835-8.

Mannuzza S, Klein RG, Konig PH, Giampino TL. Hyperactive boys almost grown up. IV. Criminality and its relationship to psychiatric status. Archives of general psychiatry 1989; 46: 1073-1079.

Masserman JH, Wechkin S, Terris W. "Altruistic" behavior in rhesus monkeys. American Journal of Psychiatry 1964; 121.

McAlonan K, Brown VJ. Orbital prefrontal cortex mediates reversal learning and not attentional set shifting in the rat. Behavioral Brain Research 2003; 146: 97-103.

McGaugh JL. Memory consolidation and the amygdala: A systems perspective. Trends in Neuroscience 2002; 25: 456-61.

McIntyre CK, Hatfield T, McGaugh JL. Amygdala norepinephrine levels after training predict inhibitory avoidance retention performance in rats. European Journal of Neuroscience 2002; 16: 1223-6.

McNaughton N, Gray JA. Anxiolytic action on the behavioural inhibition system implies multiple types of arousal contribute to anxiety. Journal of Affective Disorders 2000; 61: 161-76.

Mealey L. The sociobiology of sociopathy: An integrated evolutionary model. Behavioral and Brain Sciences 1995; 18: 523-599.

Menard J, Treit D. Effects of centrally administered anxiolytic compounds in animal models of anxiety. Neuroscience and Biobehavioral Reviews 1999; 23: 591-613.

Menon V, Adleman NE, White CD, Glover GH, Reiss AL. Error-related brain activation during a Go/NoGo response inhibition task. Human Brain Mapping 2001; 12: 131-43.

Messinger A, Squire LR, Zola SM, Albright TD. Neuronal representations of stimulus associations develop in the temporal lobe during learning. Proceedings of the National Academy of Sciences of the United States of America 2001; 98: 12239-44.

Meunier M, Bachevalier J, Mishkin M. Effects of orbital frontal and anterior cingulate lesions on object and spatial memory in rhesus monkeys. Neuropsychologia 1997; 35: 999-1015.

Milich R, Hartung CM, Martin CA, Haigler ED. Behavioral disinhibition and underlying processes in adolescents with disruptive behavior disorders. In: Routh DK, editor. Disruptive Behavior Disorders. New York: Plenum Press, 1994.

Mitchell DG, Colledge E, Leonard A, Blair RJ. Risky decisions and response reversal: is there evidence of orbitofrontal cortex dysfunction in psychopathic individuals? Neuropsychologia 2002; 40: 2013-2022.

Moffitt TE. Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. Psychological Review 1993a; 100: 674-701.

Moffitt TE. The neuropsychology of conduct disorder. Development and psychopathology 1993b; 5: 135-152.

Moffitt TE, Caspi A, Harrington H, Milne BJ. Males on the life-course-persistent and adolescence-limited antisocial pathways: follow-up at age 26 years. Development and Psychopathology 2002; 14: 179-207.

Monchi O, Petrides M, Petre V, Worsley K, Dagher A. Wisconsin Card Sorting revisited: distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. Journal of Neuroscience 2001; 21: 7733-41.

Morgan AB, Lilienfield SO. A meta-analytic review of the relation between antisocial behavior and neuropsychological measures of executive function. Clinical Psychology Review 2000; 20: 113-136.

Morris JS, Frith CD, Perrett DI, Rowland D, Young AW, Calder AJ, et al. A differential response in the human amygdala to fearful and happy facial expressions. Nature 1996; 383: 812-815.

Murphy FC, Michael A, Robbins TW, Sahakian BJ. Neuropsychological impairment in patients with major depressive disorder: the effects of feedback on task performance. Psychological Medicine 2003; 33: 455-67.

Murphy FC, Smith KA, Cowen PJ, Robbins TW, Sahakian BJ. The effects of tryptophan depletion on cognitive and affective processing in healthy volunteers. Psychopharmacology (Berl) 2002; 163: 42-53.

Murphy P. Inhibitory control in adults with Attention-Deficit/Hyperactivity Disorder. Journal of Attention Disorders 2002; 6: 1-4.

Nagahama Y, Okada T, Katsumi Y, Hayashi T, Yamauchi H, Oyanagi C, et al. Dissociable mechanisms of attentional control within the human prefrontal cortex. Cerebral Cortex 2001; 11: 85-92.

Newman JP. Psychopathic behaviour: An information processing perspective. In: Cooke DJ, Forth AE and Hare RD, editors. Psychopathy: Theory, Research and Implications for Society. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1998: 81-105.

Newman JP, Kosson DS. Passive avoidance learning in psychopathic and nonpsychopathic offenders. Journal of Abnormal Psychology 1986; 95: 252-6.

Newman JP, Patterson CM, Howland EW, Nichols SL. Passive avoidance in psychopaths: The effects of reward. Personality and Individual Differences 1990; 11: 1101-1114.

Newman JP, Patterson CM, Kosson DS. Response perseveration in psychopaths. Journal of Abnormal Psychology 1987; 96: 145-8.

Newman JP, Schmitt WA. Passive avoidance in psychopathic offenders: a replication and extension. Journal of Abnormal Psychology 1998; 107: 527-532.

Newman JP, Schmitt WA, Voss WD. The impact of motivationally neutral cues on psychopathic individuals: Assessing the generality of the response modulation hypothesis. Journal of Abnormal Psychology 1997; 106: 563-575.

Newman JP, Widom CS, Nathan S. Passive avoidance in syndromes of disinhibition: psychopathy and extraversion. Journal of Personality and Social Psychology 1985; 48: 1316-27.

Nigg JT, Blaskey LG, Huang-Pollock CL, Rappley MD. Neuropsychological executive functions and DSM-IV ADHD subtypes. Journal of the American Academy of Child and Adolescent Psychiatry 2002; 41: 59-66.

Nucci LP, Herman S. Behavioral disordered children's conceptions of moral, conventional, and personal issues. Journal of Abnormal Child Psychology 1982; 10: 411-425.

O'Brien BS, Frick PJ. Reward dominance: Associations with anxiety, conduct problems, and psychopathy in children. Journal of Abnormal Child Psychology 1996; 24: 223-240.

O'Doherty J. Can't learn without you: predictive value coding in orbitofrontal cortex requires the basolateral amygdala. Neuron 2003; 39: 731-3.

O'Doherty J, Critchley H, Deichmann R, Dolan RJ. Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. Journal of Neuroscience 2003a; 23: 7931-9.

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 2004; 304: 452-4.

O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C. Abstract reward and punishment representations in the human orbitofrontal cortex. Nature Neuroscience 2001; 4: 95-102.

O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal difference models and reward-related learning in the human brain. Neuron 2003b; 38: 329-37.

Ogloff JR, Wong S. Electrodermal and cardiovascular evidence of a coping response in psychopaths. Criminal Justice and Behaviour 1990; 17: 231-245.

Oosterlaan J, Logan GD, Sergeant JA. Response inhibition in AD/HD, CD, comorbid AD/HD + CD, anxious, and control children: a meta-analysis of studies with the stop task. Journal of Child Psychology and Psychiatry 1998; 39: 411-25.

Oosterlaan J, Scheres A, Sergeant JA. Which executive functioning deficits are associated with AD/HD, ODD/CD and comorbid AD/HD+ODD/CD? Journal of Abnormal Child Psychology 2005; 33: 69-85.

Oosterlaan J, Sergeant JA. Effects of reward and response cost on response inhibition in AD/HD, disruptive, anxious, and normal children. Journal of Abnormal Child Psychology 1998a; 26: 161-74.

Oosterlaan J, Sergeant JA. Response inhibition and response re-engagement in attention-deficit/hyperactivity disorder, disruptive, anxious and normal children. Behavioral Brain Research 1998b; 94: 33-43.

Overman WH. Sex differences in early childhood, adolescence, and adulthood on cognitive tasks that rely on orbital prefrontal cortex. Brain and Cognition 2004; 55: 134-47.

Overman WH, Bachevalier J, Schuhmann E, Ryan P. Cognitive gender differences in very young children parallel biologically based cognitive gender differences in monkeys. Behavioral Neuroscience 1996; 110: 673-84.

Packard MG. Glutamate infused posttraining into the hippocampus or caudate-putamen differentially strengthens place and response learning. Proceedings of the National Academy of the Sciences United States of America 1999; 96: 12881-6.

Packard MG, McGaugh JL. Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. Neurobiology of Learning and Memory 1996; 65: 65-72.

Pagnoni G, Zink CF, Montague PR, Berns GS. Activity in human ventral striatum locked to errors of reward prediction. Nature Neuroscience 2002; 5: 97-8.

Panksepp J. The neurobiology of emotions: Of animal brains and human feelings. In: Wagner H and Manstead A, editors. Handbook of Social Psychophysiology. Chichester: Wiley & Sons, 1989.

Panksepp J. Affective neuroscience: The foundations of human and animal emotions. New York: Oxford University Press, 1998.

Pastor MC, Molto J, Vila J, Lang PJ. Startle reflex modulation, affective ratings and autonomic reactivity in incarcerated Spanish psychopaths. Psychophysiology 2003; 40: 934-8.

Patrick CJ. Emotion and psychopathy: startling new insights. Psychophysiology 1994; 31: 319-30.

Patrick CJ, Bradley MM, Lang PJ. Emotion in the criminal psychopath: Startle reflex modulation. Journal of Abnormal Psychology 1993; 102: 82-92.

Patrick CJ, Cuthbert BN, Lang PJ. Emotion in the criminal psychopath: Fear image processing. Journal of Abnormal Psychology 1994; 103: 523-534.

Patterson CM, Newman JP. Reflectivity and learning from aversive events: toward a psychological mechanism for the syndromes of disinhibition. Psychological Review 1993; 100: 716-36.

Pennington BF, Bennetto L. Main effects or transaction in the neuropsychology of conduct disorder? Commentary on "The neuropsychology of conduct disorder". Development and Psychopathology 1993; 5: 153-164.

Pennington BF, Ozonoff S. Executive functions and Development and Psychopathology. Journal of Child Psychology and Psychiatry 1996; 37: 51-87.

Perry DG, Perry LC. Denial of suffering in the victim as a stimulus to violence in aggressive boys. Child Development 1974; 45: 55-62.

Peschardt KS, Leonard A, Morton J, Blair RJR. Differential stimulus-reward and stimulus-punishment learning in individuals with psychopathy. submitted.

Peschardt KS, Newman C, Mitchell DG, Richell RA, Leonard A, Morton J, et al. Differentiating Among Prefrontal Substrates in Psychopathy: Neuropsychological Test Findings. under revision;1.

Peterson BS, Kane MJ, Alexander GM, Lacadie C, Skudlarski P, Leung HC, et al. An event-related functional MRI study comparing interference effects in the Simon and Stroop tasks. Cognitive Brain Research 2002; 13: 427-40.

Phillips ML, Young AW, Scott SK, Calder AJ, Andrew C, Giampietro V, et al. Neural responses to facial and vocal expressions of fear and disgust. Proceedings of the Royal Society of London. Series B, Biological Sciences 1998; 265: 1809-17.

Phillips ML, Young AW, Senior C, Brammer M, Andrews C, Calder AJ, et al. A specified neural substrate for perceiving facial expressions of disgust. Nature 1997; 389: 495-498.

Pichot P. Psychopathic behavior: Approaches to research. In: Hare RD and Schalling DS, editors. Psychopathic behavior: a Historcal review - reverse. Chichester: Wiley, 1978.

Prather MD, Lavenex P, Mauldin-Jourdain ML, Mason WA, Capitanio JP, Mendoza SP, et al. Increased social fear and decreased fear of objects in monkeys with neonatal amygdala lesions. Neuroscience 2001; 106: 653-8.

Rahman S, Sahakian BJ, Hodges JR, Rogers RD, Robbins TW. Specific cognitive deficits in mild frontal variant frontotemporal dementia. Brain 1999; 122 (Pt 8): 1469-93.

Raine A. The psychopathology of crime. New Yory: Academic Press, 1997.

Raine A. Annotation: the role of prefrontal deficits, low autonomic arousal, and early health factors in the development of antisocial and aggressive behavior in children. Journal of Child Psychology and Psychiatry 2002a; 43: 417-34.

Raine A. Biosocial studies of antisocial and violent behavior in children and adults: a review. Journal of Abnormal Child Psychology 2002b; 30: 311-26.

Raven JC. Advanced Progressive Matrices, Set I. Oxford: Oxford Psychologists Press, 1976.

Reeve WV, Schandler SL. Frontal lobe functioning in adolescents with attention deficit hyperactivity disorder. Adolescence 2001; 36: 749-65.

Remijnse PL, Nielen MM, Uylings HB, Veltman DJ. Neural correlates of a reversal learning task with an affectively neutral baseline: An event-related fMRI study. Neuroimage 2005; 26: 609-18.

Rescorla RA. Response-outcome associations remain functional through interference treatments. Animal Learning and Behaviour 1996; 24: 450-458.

Rice GE. Aiding responses in rats: Not in guinea pigs. Proceedings of the Annual Convention of the American Psychological Association 1965: 105-106.

Rice GE, Gainer P. "Altruism" in the albino rat. Journal of Comparative & Physiological Psychology 1962; 55: 123-125.

Roberts AC, Robbins TW, Weiskrantz L. The Prefrontal Cortex and Cognitive Functions. Oxford: Oxford University Press, 1998.

Robins L. Sturdy childhood predictors of adult antisocial behaviour: replications from longitudinal studies. Psychological Medicine 1978; 8: 611-22.

Rogers RD, Andrews TC, Grasby PM, Brooks DJ, Robbins TW. Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans. Journal of Cognitive Neuroscience 2000; 12: 142-62.

Rogers RD, Blackshaw AJ, Middleton HC, Matthews K, Hawtin K, Crowley C, et al. Tryptophan depletion impairs stimulus-reward learning while methylphenidate disrupts attentional control in healthy young adults: implications for the monoaminergic basis of impulsive behaviour. Psychopharmacology (Berl) 1999a; 146: 482-91.

Rogers RD, Everitt BJ, Baldacchino A, Blackshaw AJ, Swainson R, Wynne K, et al. Dissociable deficits in the decision-making cognition of chronic amphetamine abusers, opiate abusers, patients with focal damage to prefrontal cortex, and tryptophan-depleted normal volunteers: evidence for monoaminergic mechanisms. Neuropsychopharmacology 1999b; 20: 322-39.

Rogers RD, Lancaster M, Wakeley J, Bhagwager Z. The effects of beta-adrenoceptor blockade on components of human decision-making. Psychopharmacology (Berl) 2004a; 172(2): 157-164.

Rogers RD, Ramnani N, Mackay C, Wilson JL, Jezzard P, Carter CS, et al. Distinct portions of anterior cingulate cortex and medial prefrontal cortex are activated by reward processing in separable phases of decision-making cognition. Biological Psychiatry 2004b; 55: 594-602.

Rolls ET. The orbitofrontal cortex. Philosophical Transactions of the Royal Society, Series B: Biological Sciences 1997; 351: 1433-1443.

Rolls ET. The Brain and Emotion. Oxford: OUP, 1999.

Rolls ET. The orbitofrontal cortex and reward. Cerebral Cortex 2000; 10: 284-294.

Rolls ET, Hornak J, Wade D, McGrath J. Emotion-related learning in patients with social and emotional changes associated with frontal lobe damage. Journal of Neurology Neurosurgery and Psychiatry 1994; 57: 1518-24.

Roussy S, Toupin J. Behavioral inhibition deficits in juvenile psychopaths. Aggressive Behavior 2000; 26: 413-424.

Rubia K, Oosterlaan J, Sergeant JA, Brandeis D, v Leeuwen T. Inhibitory dysfunction in hyperactive boys. Behavioral Brain Research 1998; 94: 25-32.

Rubia K, Smith A, Brammer M, Taylor E. Right inferior prefrontal cortex mediates response inhibition while mesial prefrontal cortex is responsible for error detection. NeuroImage 2003; 20: 351-358.

Ruff CC, Woodward TS, Laurens KR, Liddle PF. The role of the anterior cingulate cortex in conflict processing: evidence from reverse stroop interference. Neuroimage 2001; 14: 1150-8.

Rutter M, Giller H, Hagell A. Antisocial Behavior by Young People. New York, NY: Cambridge University Press, 1998.

Sandberg K, Sanberg PR, Hanin I, Fisher A, Coyle JT. Cholinergic lesion of the striatum impairs acquisition and retention of a passive avoidance response. Behavioral Neuroscience 1984; 98: 162-5.

Scerbo A, Raine A, O'Brien M, Chan CJ, Rhee C, Smiley N. Reward dominance and passive avoidance learning in adolescent psychopaths. Journal of Abnormal Child Psychology 1990; 18: 451-63.

Schall JD, Stuphorn V, Brown JW. Monitoring and control of action by the frontal lobes. Neuron 2002; 36: 309-22.

Schneider F, Gur RC, Gur RE, Muenz LR. Standardized mood induction with happy and sad facial expression. Psychiatry Research 1994; 51: 19-31.

Schoenbaum G, Chiba AA, Gallagher M. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. Nature Neuroscience 1998; 1: 155-159.

Schoenbaum G, Chiba AA, Gallagher M. Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. Journal of Neuroscience 1999; 19: 1876-84.

Schoenbaum G, Nugent SL, Saddoris MP, Setlow B. Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations. Neuroreport 2002; 13: 885-90.

Schoenbaum G, Setlow B, Saddoris MP, Gallagher M. Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. Neuron 2003; 39: 855-67.

Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science 1997; 275: 1593-9.

Scott S, Knapp M, Henderson J, Maughan B. Financial cost of social exclusion: follow up study of antisocial children into adulthood. British Medical Journal 2001; 323: 191.

Serin RC, Amos NL. The role of psychopathy in the assessment of dangerousness. International Journal of Law and Psychiatry 1995; 18: 231-238.

Setlow B, Schoenbaum G, Gallagher M. Neural encoding in ventral striatum during olfactory discrimination learning. Neuron 2003; 38: 625-36.

Seymour B, O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, Dolan RJ, et al. Temporal difference models describe higher-order learning in humans. Nature 2004; 429: 664-7.

Shallice T. From Neuropsychology to Mental Structure. Cambridge: Cambridge University Press, 1988.

Simmons J. Crime in England and Wales 2001/2002. London: Home Office, 2002.

Simonoff E, Elander J, Holmshaw J, Pickles A, Murray R, Rutter M. Predictors of antisocial personality. Continuities from childhood to adult life. British Journal of Psychiatry 2004; 184: 118-27.

Smetana JG. Preschool children's conceptions of moral and social rules. Child Development 1981; 52: 1333-1336.

Smetana JG. Preschool children's conceptions of transgressions: The effects of varying moral and conventional domain-related attributes. Developmental Psychology 1985; 21: 18-29.

Smetana JG. Understanding of social rules. In: Bennett M, editor. The child as psychologist: An introduction to the development of social cognition. New York: Harvester Wheatsheaf, 1993: 111-141.

Smetana JG, Braeges JL. The development of toddlers' moral and conventional judgments. Merrill-Palmer Quarterly: Journal of Developmental Psychology 1990; 36: 329-346.

Smith EE, Jonides J. Storage and executive processes in the frontal lobes. Science 1999; 283: 1657-1661.

Smith SS, Arnett PA, Newman JP. Neuropsychological differentiation of psychopathic and nonpsychopathic criminal offenders. Personality and Individual Differences. 1992; 13: 1233-1243.

Smith SS, Newman JP. Alcohol and drug abuse-dependence disorders in psychopathic and nonpsychopathic criminal offenders. Journal of Abnormal Psychology 1990; 99: 430-9.

Snodgrass JG, Vanderwart M. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. Journal of Experimental Psychology: Human Learning and Memory 1980; 6: 174-215.

Stevens D, Charman T, Blair RJR. Recognition of emotion in facial expressions and vocal tones in children with psychopathic tendencies. Journal of Genetic Psychology 2001; 162: 201-211.

Stuss D, T., Benson DF. The frontal lobes. New York: Raven Press, 1986.

Sustrik R, Coupland N, Blair RJR. Noradrenergic drugs and emotion recognition. in preparation.

Sutker PB. Vicarious conditioning and sociopathy. Journal of Abnormal Psychology 1970; 76: 380-386.

Sutton RS, Barto AG. Toward a modern theory of adaptive networks: expectation and prediction. Psychological Review 1981; 88: 135-70.

Swainson R, Rogers RD, Sahakian BJ, Summers BA, Polkey CE, Robbins TW. Probabilistic learning and reversal deficits in patients with Parkinson's disease or frontal or temporal lobe lesions: possible adverse effects of dopaminergic medication. Neuropsychologia 2000; 38: 596-612.

Swanson JM, Posner MI, Cantwell D, Wigal S, Crinella F, Filipek P, et al. Attention-deficit/hyperactivity disorder: Symptom domains, cognitive processes and neural networks. In: Parasuraman R, editor. The attentive brain. Massachusetts: MIT Press, 1998.

Taghzouti K, Louilot A, Herman JP, Le Moal M, Simon H. Alternation behavior, spatial discrimination, and reversal disturbances following 6-hydroxydopamine lesions in the nucleus accumbens of the rat. Behavioral and Neural Biology 1985; 44: 354-63.

Talairach J, Tournoux P. Co-planar stereotaxic atlas of the human brain. Stuttgart: Thieme, 1988.

Taylor EA, Schachar R, Thorley G, Wieselberg M. Conduct disorder and hyperactivity: I. Separation of hyperactivity and antisocial conduct in British child psychiatric patients. British Journal of Psychiatry 1986; 149: 760-767.

Thornquist MH, Zuckerman M. Psychopathy, passive-avoidance learning and basic dimensions of personality. Personality & Individual Differences 1995; 19: 525-534.

Toupin J, Dery M, Pauze R, Mercier H, Fortin L. Cognitive and familial contributions to conduct disorder in children. Journal of Child Psychology and Psychiatry 2000; 41: 333-44.

Trasler G. Relations between psychopathy and persistent criminality - methodological and theoretical issues. In: Hare RD and Schalling D, editors. Psychopathic behaviour: Approaches to research. Chichester, England: Wiley, 1978.

Trasler GB. Criminal behaviour. In: Eysenck HJ, editor. Handbook of abnormal psychology. London: Pitman, 1973.

Treit D, Menard J. Dissociations among the anxiolytic effects of septal, hippocampal, and amygdaloid lesions. Behavioral Neuroscience 1997; 111: 653-8.

Tremblay L, Schultz W. Relative reward preference in primate orbitofrontal cortex. Nature 1999; 398: 704-8.

Tremblay L, Schultz W. Reward-related neuronal activity during go-nogo task performance in primate orbitofrontal cortex. Journal of Neurophysiology 2000; 83: 1864-76.

Tremblay RE, Pihl RO, Vitaro F, Dobkin PL. Predicting early onset of male antisocial behavior from preschool behavior. Archives of General Psychiatry 1994; 51: 732–739.

Trommer BL, Hoeppner JA, Lorber R, Armstrong KJ. The go-no-go paradigm in attention deficit disorder. Annals of Neurology 1988; 24: 610-4.

Turiel E. The development of social knowledge: Morality and Convention. Cambridge: Cambridge University Press, 1983.

Viding E, Blair RJ, Moffitt TE, Plomin R. Evidence for substantial genetic risk for psychopathy in 7-year-olds. Journal of Child Psychology and Psychiatry 2005; 46: 592-7.

Vitiello B, Stoff DM. Subtypes of aggression and their relevance to child psychiatry. Journal of the American Academy of Child and Adolescent Psychiatry 1997; 36: 307-15.

Vollm BA, de Araujo IE, Cowen PJ, Rolls ET, Kringelbach ML, Smith KA, et al. Methamphetamine activates reward circuitry in drug naive human subjects. Neuropsychopharmacology 2004; 29: 1715-22.

Ward D, B. Simultaneous Inference For FMRI Data: http://afni.nimh.nih.gov/pub/dist/doc/manuals/AlphaSim.pdf, 2000.

Watanabe J, Sugiura M, Sato K, Sato Y, Maeda Y, Matsue Y, et al. The human prefrontal and parietal association cortices are involved in NO-GO

performances: an event-related fMRI study. Neuroimage 2002; 17: 1207-16.

Whalen PJ, McInerney SC, McNally RJ, Wilhelm S, Jenike MA, Rauch SL. The emotional counting Stroop paradigm: a functional magnetic resonance imaging probe of the anterior cingulate affective division. Biological Psychiatry 1998; 44: 1219-1228.

Williams D, Stott CM, Goodyer IM, Sahakian BJ. Specific language impairment with or without hyperactivity: neuropsychological evidence for frontostriatal dysfunction. Developmental Medicine and Child Neurology 2000; 42: 368-75.

Williamson S, Hare RD, Wong S. Violence: Criminal psychopaths and their victims. Canadian Journal of Behavioral Science 1987; 19: 454-462.

Williamson S, Harpur TJ, Hare RD. Abnormal processing of affective words by psychopaths. Psychophysiology 1991; 28: 260-273.

Wodushek TR, Neumann CS. Inhibitory capacity in adults with symptoms of Attention Deficit/Hyperactivity Disorder (ADHD). Archives of Clinical Neuropsychology 2003; 18: 317-30.

Wolfgang ME, Figlio R, Sellin T. Delinquency in a birth cohort. Chicago: University of Chicago Press., 1972.

Woodworth M, Porter S. In cold blood: characteristics of criminal homicides as a function of psychopathy. Journal of Abnormal Psychology 2002; 111: 436-45.